INVESTIGATIONS OF DARK MATTER USING COSMOLOGICAL SIMULATIONS

by

Lin Yang

A dissertation submitted to The Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy.

Baltimore, Maryland

May, 2017

Abstract

In the successful concordance model of cosmology, dark matter is crucial for structures to form as we observe it in the universe. Despite the overwhelming observational evidence for its existence, it is not yet directly detected, and its nature is largely unknown. Physicists propose various dark matter candidates, with masses ranging over dozens of orders of magnitude. However, both indirect and direct detection experiments for dark matter have reported no convincing results. Dark matter research is therefore critically relying on computer simulations. Using supercomputer numerical simulations, we can test the correctness of the current cosmological model, as well as obtain guidance for future detection experiments. In this dissertation, we study dark matter from several perspectives using cosmological simulations: its possible radiation, its warmth, and other related issues.

A commonly accepted candidate for dark matter is the weakly interacting massive particle (WIMP). WIMPs interact with normal matter only through the weak force (as well as gravity). It is thus extremely challenging to detect these particles directly.

ABSTRACT

However, depending on the type of dark matter, they can annihilate with other dark matter particles, or decay into high-energy photons (i.e., γ-ray). We studied the spatial distribution of possible emission components from dark matter annihilation or decay in a large simulation of a galaxy like the Milky Way. The predicted emission components can be used as templates for observations such as those from the *Fermi*/LAT γ-ray instrument, to constrain for the physical properties of dark matter.

Structure formation theory suggests that dark matter is "cold", i.e., moving non-relativistically during structure formation. However, cold dark matter predicts many more dark-matter satellites, or subhaloes, around galaxies such as the Milky Way than observed. One well-established mechanism to bring the theory in line with observations is that many of these satellites are not visible because they are too small for baryons to form stars in them. Another way is to attenuate the small-scale structure directly, positing "warm" dark matter. Using simulation, we propose a method of testing this possibility in a complementary environment, by measuring the density profile of cosmic voids. Our results suggest that there are sufficient differences between warm and cold dark matter to test using future observations.

Furthermore, our data analyzing methods are based on sophisticated data stream algorithms and newly developed Graphic Process Unit (GPU) hardware. These tools lead to other studies of dark matter as well. For example, we studied the spin alignment of dark matter halos with its environment. We show that the spin alignments

are highly related to the hierarchical levels of the cosmic web, in which the halo is located. We also studied the responses in different density variables to "ringing" the initial density field at different spatial frequencies (i.e. putting spikes in the power spectrum at a particular scale). The conventional wisdom is that power generally migrates from large to small comoving scales from the initial to final conditions. But in this work, we found that this conventional wisdom is only true for a density variable emphasizing dense regions, such as the usual overdensity field. In the log-density field, however, power stays about at the same scale but broadens. In the reciprocal-density field, emphasizing low-density regions, power moves to larger scales. This is an example of voids as "cosmic magnifying glasses." The GPU density-estimation technique was crucial for this study, allowing the density to be estimated accurately even when modestly sampled with particles. Our results provide guidance for designing future statistic analytics for dark matter and the large-scale structure of the Universe in general.

Acknowledgments

Foremost, I am sincerely grateful to my adviser Prof. Alex Szalay. He is the most extraordinary physicist/computer-scientist I have ever met. His enthusiasm of applying the state-of-art technology to solve the most challenging problems in today's Astrophysical research is always inspiring me. I am thankful for his broad guidance to both my research and my future goals. I am deeply indebted to the free environment he has created for me – no matter what I am pursuing, I do not need to worry about my funding source. I am also thankful for his introductions of great collaborators to me. Without these collaborators, this dissertation would not be completed.

I would like to express my gratitude to my dissertation committee, Prof. Rosemary Wyse, Prof. Collin Broholm, Prof. Vladamir Braverman and Prof. Tamas Budavari.

Moreover, I would like to thank my collaborators, Joe Silk, Rosie Wyse, Mike Kuhlen, Mark Neyrinck, Miguel Aragon and Liang Dai. I am grateful for all the help, guidance and insightful discussions.

ACKNOWLEDGMENTS

I have spent nearly six memorable years at Johns Hopkins University with my friends, Liang Dai, Yaofu Zhou, Lei Feng, Can You, Zaoxin Liu, Tuo Zhao, Steve Chestnut, Nikita Ivkin and many others. I cannot imagine how I could survive without the companion of these friends.

I owe a great deal to my parents for their understanding and selfless support from the other side of the Earth. I have never faced a financial crisis that is usually faced by people who have a similar family background like me. They are standing deep in the root of my life – no matter where I am going, I will never feel alone.

Last but not least, I would like to express gratefulness to my beloved wife, Lizi Xie for her love and care, understanding and support. We have met since we were high school students. I cannot imagine how much she has sacrificed for leaving behind the colorful life in China to be together with me. We have shared together every moment of our lives. Words are not enough to express my gratitude for her unconditional love and support.

Dedication

To my mother and father, Yanzhen Sheng & Zhongfa Yang.

To my beloved wife, Lizi Xie.

Contents

Abstract **ii**

Acknowledgments **v**

List of Figures **xii**

1 Introduction **1**

2 The Radiation of Dark Matter **11**

2.1 Dark Matter Annihilation And Decay 16

2.2 Simulation Data and Map Production Algorithm 18

2.3 Corrections 23

2.3.1 Sommerfeld Enhancement 23

2.3.2 P-wave Annhiliation 24

2.3.3 Baryonic Matter and Adiabatic Contraction 25

2.4 All-Sky Maps and Gamma-Ray Signals from Different Sky Regions 26

2.5 Discussions 32

3 Shallower Voids with Warm Dark Matter **35**

3.1 Simulations 38
3.2 The Cosmic Web in a WDM Scenario 42
3.3 Void Detection and Properties 48
3.4 Density profiles 54
3.5 Conclusion and Discussion 57

4 Response of Power Spectra to Initial Spikes in CDM Regime 59

4.1 Methods 62
4.2 Results 65
4.3 Toy model for power spreading 70
4.4 Covariance Matrices and Information 73
4.5 Conclusions 76

5 Dark Matter Halo Spin Alignment Under CDM 78

5.1 N-body Simulation and DM haloes: the MIP correlated ensemble 80
5.1.1 Halos 82
5.2 MMF-2: Hierarchical Identification of Filaments and Walls 83
5.2.1 Hierarchical space 84
5.2.2 Density fields 85
5.2.3 The filament/sub-filament sample 86
5.2.4 Spin alignment in filaments and walls 87
5.3 Results 87
5.4 Conclusions and discussion 88
5.4.1 On the origin and fate of the spin alignment 90

6 Simulation Data Analyzing Methods **92**

6.1 Dark Matter Annihilation Calculation 94

6.1.1 GPU Based STR Projection (GSP) Method 95

6.1.2 Comparasion 97

6.1.3 Potential Application 98

6.1.4 Conclusion 99

6.2 Lagrangian Tessellation Field Estimation 99

6.3 Streaming Algorithm Speeds up Halo Finding 102

6.3.1 Streaming Algorithm 106

6.3.2 Streaming Data Model 106

Definitions 106

Heavy Hitter 107

Data Transformation 108

Heavy Hitter and Dense Cells 109

6.3.3 Streaming Algorithms for Heavy Hitter Problem 110

The Count-Sketch Algorithm 111

The Pick-and-Drop Sampling Algorithm 112

6.3.4 Implementation 113

Simulation Data 113

Implementation Details 115

Pick-and-Drop-based Halo Finder 115

Shifting Method 118

6.3.5 Evaluation 119
Correctness 121
Stability 124
Memory Usage 125
6.3.6 Conclusion 127

Bibliography **129**

Vita **163**

List of Figures

1.1 The illustration of a cosmic web. The box size is 100Mpc/h. The color encodes the density of the dark matter. Blue region means less dense. . . . 4

2.1 Mollweide projection of DM annihilation flux. *From (a) to (f)*: pure annihilation, annihilation with Sommerfeld enhancement $1/v$ correction, annihilation with Sommerfeld enhancement $1/v^2$ correction, annihilation with v^2 correction, annihilation with adiabatic contraction, pure decay. A Gaussian filter with FWHM 0.5° is applied. The flux is normalized such that the host halo has flux unity. 16

2.2 The density profile of the host halo and contracted profile. The solid black line is the original density profile of the host halo. After contraction, the profile becomes the long dashed line. The dotted line is the NFW profile with parameters given by Ref.,[1] with $\gamma = 1.24, \alpha = 1$ and $\beta = 1.76$. The dotted dash line is the NFW$_C$ profile describing the adiabatic contraction with $\gamma = 1.37, \alpha = 0.76$ and $\beta = 3.3$. The normalization is chosen such that the intersection point of the contracted and non-contracted profiles are at a density of 1. See text for details. 21

2.3 The flux annular profile of different cases. *Left panel*: the angular profile of $0 - 90^\circ$ region of the sky maps; *right panel*: the profile of the zoomed in region of the left panel in the inner $0 - 5^\circ$. For legends of lines: solid, pure annihilation; thin dash dotted, annihilation with Sommerfeld enhancement $1/v$ correction; thick dash dotted, annihilation with Sommerfeld enhancement $1/v^2$ correction; dotted, annihilation with v^2 correction; thick dashed, annihilation with adiabatic contraction; thin dashed, pure decay. The flux are normalized such that the host halo has flux unity. See text for details. 22

2.4 Angular profile of different sky regions. *Upper left panel*: the longitudinal profile of $|b| > 8^\circ$; *upper right panel*: the longitudinal profile of $|b| \leq 5^\circ$; *lower left panel*: latitude profile of $|l| \leq 30^\circ$ and *lower right panel*: latitude profile of $|l| \geq 90^\circ$. For legends of lines: black solid, the DGE model ${}^{\rm S}{\rm S}^{\rm Z}4^{\rm R}20^{\rm T}150^{\rm C}5$ from Fermi-data; red dash dotted, pure annihilation; blue dashed, annihilation with $1/v$ correction; green dotted, annihilation with $1/v^2$ correction; solid with "x" symbols, annihilation with adiabatic contraction; magenta solid with "+" symbols, annihilation with v^2 correction; cyan solid, pure decay. All profiles are normalized so that they have maximum unity. See text for details. 27

2.5 Color map of the γ-ray flux in the inner $5^\circ \times 5^\circ$ produced by DM annihilation/decay. From (a) to (f): pure annihilation, annihilation with Sommerfeld enhancement $1/v$ correction, annihilation with Sommerfeld enhancement $1/v^2$ correction, annihilation with v^2 correction, annihilation with adiabatic contraction, pure decay. The flux are normalized such that the center pixel has flux unity. 28

2.6 The flux function for different cases. *Solid*: pure annihilation. *Thin dash-dotted*: annihilation with Sommerfeld enhancement $1/v$ correction. *Thick dash-dotted*: annihilation with Sommerfeld enhancement $1/v^2$ correction. *Dotted*: annihilation with v^2 correction. *Thick dashed*: annihilation with adiabatic contraction. *Thin dashed*: pure decay. The fluxes are normalized such that the host halo has flux unity. See text for details. 31

2.7 The latitude distribution of detected point sources. The detection threshold is chosen such that the maximum number of sources is on the order of 100. Sources are counted in latitude bins of 10°. The legend is the same as Fig. 1.6. See text for details. 32

3.1 The minimum density expected in a $(100\,h^{-1}\,\mathrm{Mpc})^3$ volume, as a function of simulation resolution. The four curves show the four cases investigated below: the WDM cut-off scale $\alpha = 0$ (CDM), 0.05, 0.1, and $0.2\,h^{-1}\,\mathrm{Mpc}$. The cell size is the initial comoving interparticle separation. A dashed line shows the cell size for the resolution used in the simulations used in this work. 41

3.2 LTFE density field slices, showing $\ln(1+\delta)$. From top left to bottom right: CDM, WDM with $\alpha = 0.05\ h^{-1}\,\mathrm{Mpc}$, WDM with $\alpha = 0.1\ h^{-1}\,\mathrm{Mpc}$and WDM with $\alpha = 0.2\,h^{-1}\,\mathrm{Mpc}$. 43

3.3 PDFs of particle densities for the four simulations. A distinctive peak arises from filament and wall particles at middling densities in a WDM scenario. The V, W, F, and H curves add up to the total: they separate out void, filament, wall, and halo particles, with crossings along 0, 1, 2, and 3 orthogonal axes. $\alpha = 0$ corresponds to CDM; as α increases, the WDM becomes warmer. In this mass-weighted PDF, each particle (Lagrangian element of initial spacing $0.2\,h^{-1}\,\mathrm{Mpc}$) enters once. The dashed magenta curve shows the expression in Eq. (2.5). 45

3.4 $z = 0$ Voronoi particle densities $\ln(1+\delta)$ on a 2D Lagrangian sheet, with one particle per $0.2\,h^{-1}\,\mathrm{Mpc}$ comoving Lagrangian pixel. Each panel shows a 2D, 256^2 slice, a quadrant of a full 512^2 slice. 46

3.5 Same as Fig. 2.4, with ORIGAMI morphologies added: black, red, and white contours separate void, wall, filament, and halo particle morphologies. . . . 47

3.6 The void radius r_{eff} distribution for different dark matter models. 50

3.7 The void core density distribution for different dark matter models. 51

3.8 The distribution of $r_{0.2}$ for different dark matter models. 52

3.9 Void density profiles measured and scaled with $r_{\rm eff}$. The origin is the density minimum of each void. The left panel shows voids in the 1-5 $h^{-1}\,$Mpc radius bin; the right panel shows results from 5-10 $h^{-1}\,$Mpc voids. Error bars show the 2σ error, dividing by $\sqrt{N}$, where N is the number of stacked voids. . . . 53

3.10 Void density profiles as in Fig. 2.9, except scaled with $r_{0.2}$. The left panel shows the voids in the $0.5 - 1.0\,h^{-1}\,$Mpc radius bin; the right panel shows the $1.0 - 3.0\,h^{-1}\,$Mpc bin. 53

3.11 Void density profiles as in Fig. 2.10, except using void volume centroids as centers. 53

4.1 Plots of the transformed density-field slices from simulations used. The width of each panel is $128\,h^{-1}\,$Mpc, half the box size, and the pixel size is $1\,h^{-1}\,$Mpc. The Lagrangian-tessellation density estimate allows accurate density estimates deep within voids, with suppressed particle-discreteness effects compared to other methods. 63

4.2 From top to bottom, $z = 0$ power spectra from initial-spike simulations of various transformed-density power spectra, along with the spiked initial conditions at $z = 127$. Power spectra are rainbow-colored according to the wavelength of their initial spike, from short (violet) to long (red). $P_{1/(1+\delta)}$ is divided by 10^3 for clarity. $P_{{\rm Gauss}(\delta)}$, investigated below, is omitted because it is almost indistinguishable from $P_{\ln(1+\delta)}$. 66

4.3 Plots of the matrix G_{ij}, defined in Eq. (3.1), showing the response of final-conditions power spectra to initial spikes. The spikes in the initial conditions are shown with dotted curves, rainbow-colored from red to violet going from low to high frequency. Corresponding power spectra of final-conditions simulations appear as solid curves. From top to bottom, the final-conditions resting place of a moderate-scale (e.g. green) spike moves from small to large scales, as each transformation increasingly emphasizes underdense regions. For clarity, power spectra from only odd-numbered spikes are shown. The 'model' is a toy model of power spreading based on a local spherical collapse or expansion of volume elements, given in Eq. (3.4). The black curves show the density propagator (CS06). 68

4.4 G_{ij}'s as in Fig. 3.3, shown in matrix form. 69

4.5 The non-Gaussian part of the δ power-spectrum covariance T_{ij}, as measured from G^{δ}_{ij} using the linear-response model of Eq. (3.6), and using an 'α model' approximation to T_{ij} found in.[2] Because of the steep increase, we use a $\sinh^{-1}$ transform for plotting, which becomes logarithmic for large values of its argument. While the α model should not be taken too seriously as it is only an approximation, the qualitative agreement (except, perhaps, far from the diagonal) between the two plots suggests that the linear-response model captures most of the relevant effects. 75

5.1 The MMF-2 method. Top panels: hierarchical space. We show the volume rendering of the density field at $z = 0$ for linear-regime smoothing at 4, 2 Mpc h^{-1} and no smoothing (left, center and rights panels respectively). The thickness of the slices was chosen to show as many structures as possible while avoiding confusion. Bottom panels: Hierarchical identification of filaments. The left and center panels show the density field inside filaments (F_1) and sub-filaments (F_2) respectively. We highlight the filament mask as a semitransparent surface. The hierarchy of filaments (blue) and sub-filaments (red) is shown in the bottom right panel. 81

5.2 Spin alignment of haloes as function of mass in filaments (left panel) and walls (right panel). The alignment is measured with the mean $\cos\theta$ where 0.5 corresponds to no preferred alignment. The sample was divided in haloes in filaments/walls (solid lines) and sub-filaments/sub-walls (dashed lines). Error bars per bin were computed from 1000 random realizations with the same number of points as their corresponding bin. 82

5.3 Distribution of angles between the halo's spin and its parent filament (black line) or sub-filament (red line). The samples were divided as high-mass haloes (solid line) and low-mass haloes (dashed line). The transition mass for filaments and sub-filaments corresponds to $\sim 1.5 \times 10^{12}$ and $\sim 4 \times 10^{11}$ $M_\odot$ h^{-1} respectively. 91

6.1 STR Projection. A circle on the sphere is projected to a circle on the projection plane. 96

6.2 Point Sprite Rendering of Annihilation Signal. Every particle's Gaussian profile is projected to a stretched Gaussian. The sphere is divided into two hemispheres to reduce the distortion. Left is the north hemisphere, right is the south hemisphere. Color: Blue: -.5 – Red: 4.5; Units: Log10[GeV^2 cm^{-6} kpc]. 96

6.3 Mollweide Projection of the All Sky Flux Maps. (a) is generated by the original slow method; (b) is generated by the GSP method. (c) shows their difference. 97

6.4 Fig-4: GSP Method Works as Virtual Observatory. Left panel simulates observing at the Galaxy center on earth with a view angle of 78 degrees (each pixel is about 9 arcmin, similar to Fermi-LATs observation). Right panel is the radial profile of the signal distribution, restoring Ref.[3]'s Fig-4. 98

6.5 A cubic cell (left) is split into 6 tetrahedra. 100

6.6 A sheet intersects with a tetrahedron. There are either two triangles or one triangle on the sheet. 101

6.7 Count-Sketch Algorithm 112

6.8 Pick-and-Drop Algorithm 113

6.9 Halo mass distribution of various halo finders. 114

6.10 Count-Sketch Algorithm 116

6.11 Pick-and-Drop Sampling 117

6.12 Halo Finder Procedure 117

6.13 Left Panel: Measures of the disagreement between PD and CS, and various in-memory algorithms. The percentage shown is the fraction of haloes farther than a half-cell diagonal ($0.5\sqrt{3}$ Mpc/h) from PD or CS halo positions. Right Panel: The number of top-1000 FoF haloes farther than a distance d away from any top-1000 halo from the algorithm of each curve. 119

6.14 Number of detected halos by our two algorithms. The solid lines correspond to (CS) and the dashed lines to (PD). The dotted line at $k = 1000$ shows our selection criteria. The x axis is the threshold in the number of particles allocated to the heavy hitter. The cyan color denotes the total number of detections, the blue curves are the true positives (TP), and the red curves are ethe false positives (FP). 122

6.15 This ROC curve shows the tradeoff between true and false detections as a function of threshold. The figure plots TPR vs FPR on a log-log scale. The two thresholds are shown with symbols, the circle denotes 1000, and the square is 900. 124

6.16 The top 1000 heavy hitters are rank-ordered by the number of their particles. We also computed a rank of the corresponding FoF halo. The linked pairs of ranks are plotted. One can see that if we adopted a cut at $k = 900$, it would eliminate a lot of the false positives. 124

6.17 Each line on the graph represents the top 1000 halo centers found with Pick-and-Drop sampling, Count-Sketch, and in-memory algorithms, as described in section 5.3.4. The comparison with FOF is shown in Fig.5.13. The shaded area (too small to be visible) shows the variation due to randomness. . . . 126

Chapter 1

Introduction

For nearly 90 years,[4] astronomers have accumulated evidence that some mysterious dark matter (DM) dominates the matter content of the universe. One of the most robust forms of evidence is the *rotation curve* (denoted as $v(r)$) of spiral galaxies. Here $v(r)$ is the typical rotational velocity of matter at a distance r to the center of the galaxy. For spiral galaxies, rotation curves can be obtained by measuring the 21cm emissions. The source of these emissions are the neutral hydrogen clouds spreading over the spiral disk. NGC 3198[5] is one of the galaxies that are measured extensively. The rotational curve of NGC 3198 is almost flat after a rise near $r = 0$. Similar curves are observed in almost all other spiral galaxies as well. If spiral galaxies are constituted entirely by normal matter, their mass should be proportional to the total luminosity. With only measurement of light emitting from NGC 3198, the expected rotation curve is inversely proportional to r, and deviates from a flat curve. Further more, the mass-to-light ratio (Υ) is estimated to be $\Upsilon \geq 30h$,[5]

where h is the dimensionless Hubble parameter ($h = 0.68 \pm 0.6$).[6] However, in a solar neighborhood, $\Upsilon \approx 5$.[7] This large discrepancy suggests that most matter in many external galaxies is unseen.

In addition to the rotational curves, the evidence for DM also emerges from greater scales. A well-known example is the light-to-mass ratio, Υ, of galaxy clusters. Using X-ray gas temperature measurements (e.g. Ref[8]) and gravitational lensing (e.g. Ref.[9]), the mass-to-light ratio is estimated roughly to be $\Upsilon \approx 330 - 620 M_{\odot}/L_{\odot}$[7] – a larger discrepancy to the one in a solar neighborhood. It is worth mentioning that Zwicky[4] published the first measurement for Coma cluster in 1933. Another stunning example is the Bullet cluster, in which two galaxy clusters are colliding. A strong separation between its gravitational lensing center and X-ray emission center is observed.[10] Since gravitational lensing center is close to the center of mass and X-ray emission follows along baryonic matter, this cluster provides arguably the best evidence of DM so far.

The precise measurement from Cosmic Microwave Background (CMB)[6] gives the evidence for DM on the largest scale ever. CMB contains the relic photons decoupled from baryonic matter around 300,000 years after the birth of the universe. The initial temperature of CMB is around 3,000K, under which atomic hydrogen starts to form. The CMB photons traveled freely through the universe as it expanded and cooled. At the present time, CMB is a cold sea of photons with an average temperature of 2.7K. The temperature of CMB is so smooth spatially that the maximum relative deviation to the mean temperature is about 10^{-4}. However, these tiny fluctuations of CMB temperature encode sufficient information of the matter components of the early universe. For example, the fluctuations

have large correlations at a certain angular radius. The phenomena correspond to the peak structures of the CMB angular power spectrum. The peaks suggest that sound waves are transmitting through the initial photon-baryon plasma. The size of the first peak determines the size of the sound horizon as well as the curvature of the universe. The alternating peak heights in the temperature power spectrum of CMB requires a second self-gravitating fluid (i.e. DM) other than the photon-baryon.

Dark matter is also required to explain the structure formation of the universe. By the cosmological principles, the universe ought to be both isotropic and homogeneous, i.e., matter distributed uniformly observing from any direction and any location. However, observations (e.g.[11]) show that the matter distribution is far more clumped – the existence of the Milky Way is a consequence of the clumpiness. The standard ΛCDM model explains why the universe is perfectly smooth at the very early age with only small quantum fluctuations. The inflation of the universe, shortly after the big bang, drastically stretched the space and time. These small fluctuations were frozen as tiny ripples among the uniform matter background. The ΛCDM model stresses that the matter of the universe is dominated by "cold" dark matter (CDM). CDM forms clumps under gravity around the seeds – the initial ripples of quantum fluctuations, and finally forms the cosmic web – the large scale structure in DM: halos, walls, and filaments (see Figure 1.1 for example). Ordinary matter, e.g. gases and dust, is clumping inside the potential wells of the cosmic web and forms stars, galaxies and galaxy clusters, the large scale structure we observe today.

ΛCDM is so successful that it matches observation from all large scales. There might be still problems that DM is not entirely "cold", i.e., moves relativistically during the epoch of

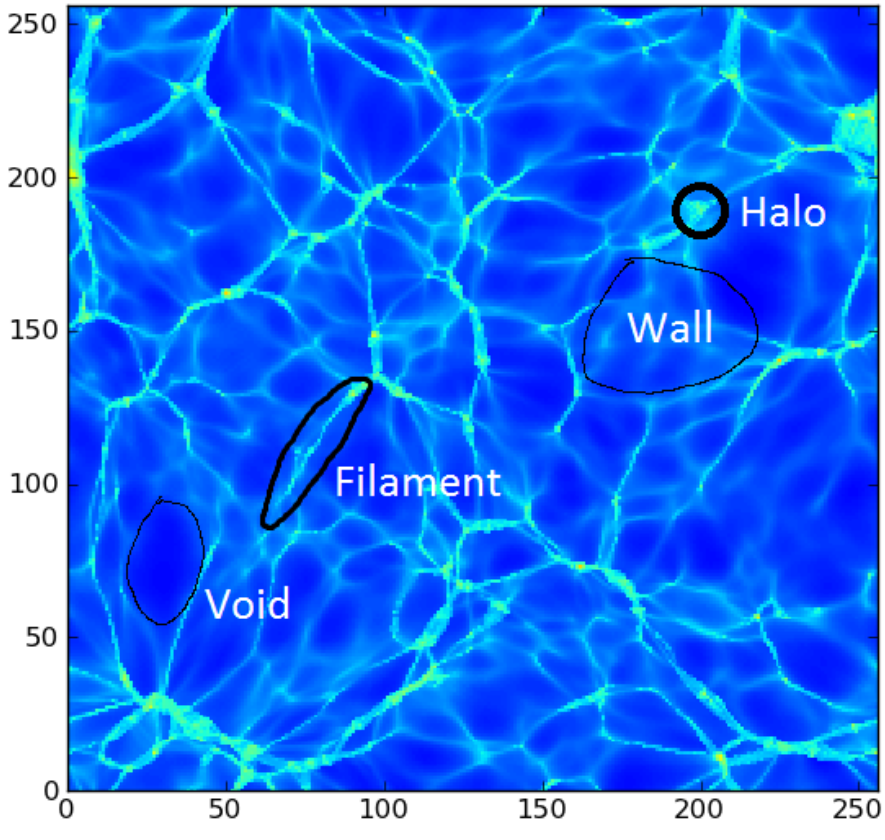

Figure 1.1: The illustration of a cosmic web. The box size is 100Mpc/h. The color encodes the density of the dark matter. Blue region means less dense.

structure formation. We will address this issue in later chapters of this thesis. Nevertheless, the evidence for DM is overwhelming.

Up until recently, suggested candidates for DM span from axions (10^{-21}eV= $10^{-88}M_{\odot}$) to massive black holes ($10^{4}M_{\odot}$). The mass of these candidates ranges over almost 100 orders of magnitude. Among those candidates, one of the most natural candidate classes is the so-called "weakly interacting massive particles" (WIMP). Since the standard cosmological model suggests that the universe was created from a big bang, it is extremely high temperate and dense in its initial state. A thermal creation scheme for DM has been proposed accordingly. In this scheme, thermal equilibrium was achieved for WIMP in the very early stage. During that period, the number density of WIMP is roughly the number density of photons. As the universe expanding, the temperature cooled down. If the thermal equilibrium were continuously maintained until today, one would expect that the number density of WIMP is $\propto \exp\left(-M_{WIMP}/T\right)$, which contributes to a tiny portion of the total number density. If WIMP is a major component of today's DM, then there is a time it"freeze-out" from equilibrium (e.g. Ref.[12]). Boltzmann equation gives an accurate prediction for the WIMP density

$$\Omega_W \approx \frac{10^{-26}\mathrm{cm}^3\mathrm{s}^{-1}}{\langle\sigma v\rangle}$$

where $\langle\sigma v\rangle$ is the thermally averaged cross section for two particles to annihilate in to standard model particles. Remarkably, according to the ΛCDM model, $\Omega_m \approx 0.3$, the above equation gives that $\langle\sigma v\rangle \approx 10^{-25}\mathrm{cm}^3\mathrm{s}^{-1}$, which is the scale of weak interaction.[12] The super symmetry model (SUSY) of particle physics provides stable candidates (e.g. neutralinos) that are both massive and weakly interacting, which suggests that WIMP is a promising

DM candidate. Other candidates including axions, primordial black hole, Massive Compact Halo Object (Macho), all have been proposed based on different physical motivations. For example, axions are generated by symmetry broken and thus do not need a thermal creation but still "cold"; massive neutrino has a small mass and provides a candidate for hot dark matter (HDM). These candidates are out of the scope of this thesis; I refer the reader to Ref.[7,13] for more detailed survey and study. There are also other explanations for the DM puzzle. For instance, the Modified Newtonian Dymanics (MOND) (e.g.[14]). This theory indeed can explain the flat rotation curve of galaxies, but it is not yet able to provide a consistent cosmological theory.

Searching for DM, especially WIMP, is now one of the major themes in experimental particle physics and particle astrophysics. Direct detection assumes WIMP particles can elastically collide with light nucleons (e.g. Xe, Ge, and Ar), causing which to recoil. Separating the WIMP recoil events from background events (e.g. all the interactions with standard model particles) enables a direct detection for WIMPs. Although possible in theory, these detections are extremely challenging to conduct due to the small cross section of WIMP-nucleon interaction. Currently, there are two leading classes of experimental design for DM detection. The first is based on solid state physics, the Cryogenic Dark Matter Search (CDMS).[15] This design uses silicon crystal in a cryogenic environment as DM detectors. The detector can read out both phonon and ionization signals using superconducting sensors and thus is capable of detection of elastic collision signal and rejection of non-elastic ones. Another experimental design is using liquefied noble gases (e.g. Xenon). These experiments are able to read out both scintillation and ionization signals. Experiments of this

design are XENON[16,17] and LUX (Large Underground Xenon[18]). Up to the writing time of this thesis, no convincing positive signal of WIMP-nucleon collision has been observed.

Other than direct searches, indirect searches are also appealing to physicists. Methods of this kind include, but not limited to, searching for WIMP annihilation signals from the galaxy and studying gravitational lensing mapping of galaxy clusters.[19] Several experiments observed signal excesses, which cannot be explained fully by known sources. For instance, recently PAMELA,[20,21] ATIC,[22] Fermi-LAT[23,24] and HESS[25] observed excesses of electron positron signals in $100 - 1000$ GeV energies. DM annihilation/decay to standard model particles may well explain these exciting results. However, none of them convincingly points toward DM because there are alternative astrophysical explanations.

Nevertheless, both direct and indirect detection of DM needs an accurate prediction of the DM density profiles, which are hardly known. Due to the high nonlinearity of gravitational interactions, even under the ΛCDM model, whose initial condition can be carefully predicted by inflation theory, analytical calculations of DM distribution of the present time is not possible. Thanks to the fast development of computational technology (e.g. Moore's law [1]), we are now capable of direct numeric calculation of the DM gravitational evolution. Under the assumption that DM is the dominating matter component, pure DM simulations have been performed in many scales of sizes. Almost all the current simulations use tracer particles to integrate the Newtonian approximation to the Einstein's field equations (i.e., Roberson-Walk metric). Simulations of this kind are called N-body simulation; namely, the simulations compute the trajectories of a vast number of tracer particles under gravity. To

[1] that the number of transistors per square inch on integrated circuits had doubled every year since their invention.

obtain more accurate simulations, we need to use a greater number of tracer particles. Thus for decades, the number of particles in simulations grows exponentially.

The first numeric simulation of DM evolution is by Press & Schechter,[26] who studied the DM halo mass functions in their seminar paper about the Press-Schechter formalism. For large scale structure formation, an incomplete list of simulations includes, Millennium Run ($10 \sim 300 \times 10^9$ particles,[27–29]), Horizon ($69 \sim 370 \times 10^9$ particles[30]) and DEUS FUR (550×10^9 particles[31]). For large scale structure formation, Millennium simulation[28] is one of the most well studied and analyzed simulations. This simulation contains 10×10^9 particles in a 500Mpc periodic box. Galactic scale simulations, on the other hand, focus on studying the formation of DM halos on the scale of a galaxy. For example, Via Lactea II simulation,[1] Aquarius simulation[32] and the GHALO run.[33] The three simulations used a similar number of particles in a Mpc-sized box. Using these simulations, we can learn a great deal of the structure of the DM halo our galaxy is embedded in. Newer and larger scaled simulations have been carrying out. These newly planned or under going simulations are focusing on more and more accurate integration of the Newtonian equations and also taken into account more physics, e.g., the physics of baryonic matter, which gives a more realistic simulation of how galaxies are formed. Numeric simulations play a vital role in understanding how structures of the universe are forming, and provide accurate predictions of the statistics of large scale surveys based on the current cosmological model.

In most of the studies in this dissertation, effects of baryonic matter are negligible. We thus use pure DM simulations to investigate the properties of DM. Details of the simulations will be presented separately in each chapter. We mainly focus on the following three aspects

of DM.

- Understanding the effects of DM spatial distribution on indirect detection. Using cosmological simulation we can predict how the annihilation/decay signal of DM is distributed spatially. We produced statistically meaningful templates to be fitted to future γ-ray observations. This study is presented in Chapter 2.

- Understanding how the warmth would affect the matter distribution in cosmic voids. Although CDM model achieves tremendous success in explaining the structure formation of the universe, on small scales (e.g. galactic scale), CDM model does match observations very well. For example, 1) the "missing satellite problem" – there are far few number of satellite galaxies observed than predicted by CDM model. 2) the halo density profile problem – the density profile of CDM model does not match that the observed profiles in the galactic center. 3) the number of dwarf galaxies expected in local voids are less than the CDM prediction. To resolve these problems, we revisit the warm dark matter (WDM) model using different sets of cosmological simulation. Under different warmth, we predict different void density profiles, which can be verified by future density estimation of cosmic voids. This study is presented in Chapter 3.

- Understanding the dynamics of CDM. Due to the high nonlinearity of the gravitational interactions, late-time structure formation, and DM evolution are far less understood from an analytical level. Using simulation, we contribute to the understanding of the kinematics of DM from two aspects. Firstly, we show that how a spike planted in the initial over-density power spectrum would evolve in the final time. We show differently

transformed over-densities can illustrate the spike migration differently, emphasizing on higher density region or lower density regions. Our results can be used to design new measurement of density statistics in theory and also in future observations. This study is presented in Chapter 4. Secondly, we show that the spin of DM halo cosmic filaments aligns to the host structure differently in different levels of the cosmic web hierarchy. This study is presented in Chapter 5.

Lastly, we presented the data analytical tools in Chapter 6. Since cosmological data set usually contains more than 10^9 particles, any statistical calculation is typically a highly non-trivial task. For example, a single calculation of the DM annihilation/decay signal might require 100 CPU hours, which might well be a supercomputer task. Our analytical tools exploit fast GPU technology as well as new progress in algorithmic research that significantly speeds up the calculation. In particular, we convert the conventional serial algorithms to parallel ones running on GPUs, which can speed up the whole calculation by nearly 1000-folds. Our analytical tools made the physics research in this dissertation possible.

Chapter 2

The Radiation of Dark Matter

This chapter is heavily based on my published article (Ref.[34]). The author list of this paper is Lin Yang, Joseph Silk, Alexander S Szalay, Rosemary FG Wyse, Brandon Bozek and Piero Madau.

Despite the overwhelming observational evidences (e.g.[13]), the physical properties of dark matter (DM) is vastly unknown. Is the dark matter a standard model particle or a super symmetric model particle? Or is it a particle that does not fit to the current framework of particle physics? To answer these questions, we need to detect the mysterious DM particle. Searching for DM becomes now a major research topic in both experimental particle physics as well as observational astrophysics. Particle physicists focus on direct detection of DM particles. However, it is well known that if DM particles interact with normal matter, the interaction shall not be stronger than weak interaction. Thus such a detection becomes extremely challenging. In fact, up to the date of writing this dissertation,

no positive results have been reported. Numerous negative results (e.g.[17,35]) suggest strong constraints on the interaction strength of DM particles with normal matter.

On the contrary, several plausible signals have been proposed as candidates of DM signals from the astrophysical observations, which are indirect methods of detecting DM. These exciting results including measurements of the end products of DM annihilation/decay (e.g. γ-ray, e^+e^- pairs[36,37]) or mapping of gravitationally lensed structures (e.g.[38]). If DM is a form of a thermal relic particle, which was once in the thermal equilibrium of the cosmic plasma in the very early universe, then the number density of DM particles can be carefully predicted using Boltzmann's equation. If this is the case, then the theory requires that two DM particles can annihilate with each other. The radiation from annihilation would then light up Galactic substructures by the resulting γ-ray emission. DM particle decay products may also be a source that is comparable to annihilation. Moreover, particle decay does not require DM as a form of a thermal relic. One such example is axion, which is generated from symmetry broken.

In this chapter, we focus on the WIMP component of dark matter particles. The absolute radiation intensity depends on the particle annihilation/decay cross section, particle type, particle mass, and astrophysical distributions, all of which are poorly known. Fortunately, the spatial *relative* annihilation/decay luminosity depends on only a few parameters, among which, particle number density is a major factor. For instance, the decay signal is proportional to density while the annihilation luminosity is proportional to density squared. Other factors affecting the relative luminosity are the thermal cross section $\langle \sigma v \rangle$, Sommerfeld boost,[39] s-wave suppressed annihilations and even baryonic suppressing of the

density profile. Sommerfeld boost is a nonperturbative increasing in the annihilation cross-section at low velocities. It is the result of a generic attractive force between the incident DM particles. Baryons may affect the density profiles of DM halos. One of such effects is the "adiabatic contraction", which is induced by the infall of baryonic matter, that could steepen the DM density profile substantially. From a numeric simulation, this effect can change the normal NFW[40] profile r^{-1}, to a profile that is proportional to $r^{-1.9}$ (e.g., Ref.[41]), where r is the radius from the center of the DM halo. Last but not least, s-wave suppressing of DM annihilation may also affect the annihilation cross-section. This effect has not been previously discussed in the Galactic context. But it might be important for spin-dependent interactions of DM particle candidates.

Diffuse Galactic γ-ray emission (DGE) is believed to be produced by interactions between cosmic rays and the interstellar medium (ISM). The launch of the Fermi Gamma-ray Satellite, with its Large Area Telescope (LAT), enabled a detailed study of cosmic-ray origin and propagation, and of the interstellar medium. In a recent study,[23] the Fermi-LAT team published an analysis of the measurements of the diffuse γ-ray emission from the first 21 months of the Fermi mission. They compared the data with models generated by the GALPROP[1] code, and showed that these emission templates underpredict the Fermi-LAT data at energies of a few GeV in the inner Galaxy. This can possibly be explained by undetected point-source populations and variations of the cosmic ray spectra. Ref.[42] has performed an analysis of the DGE in the Milky Way halo region to search for a signal from dark matter annihilations or decays. They considered a region covering the central part of the Galactic

[1] http://galprop.stanford.edu/

halo while masking out the Galactic plane. In such a region, the DM signal would have a large S/N ratio and would not depend on detailed assumptions about the center profile, e.g. the assumptions of a Navarro-Frenk-White (NFW) profile or a cored profile would only differ by a factor of ~ 2. This paper provided conservative limits on the DM cross-section assuming that all the γ-ray signal comes from DM in this region. They also provided more stringent limits based on modeling the foreground γ-ray signal with the GALPROP code. Their results impact the possible mass range that DM is produced thermally in the early universe and challenge the DM annihilation interpretation of PAMELA/Fermi-LAT cosmic ray anomalies. However, several points could be improved: 1) although using a NFW profile for modeling pure DM and a cored profile for modeling the baryonic matter effect covers extreme cases, it is still worth trying a more realistic profile directly from simulations; 2) in addition to the mass and cross section of the DM particle, the different annihilation/decay schemes may also play an important role, e.g. subhalo effects would become considerably strong in the region of interest where tidal disruption plays a role and especially when considering possible Sommerfeld enhancements; and 3) instead of detecting DM signal from the region that masks out the Galactic Center region, it may be of interest to search for a possible signal from the center or anticenter regions. Therefore, a good template for the Galactic center or the full sky should provide more constraints on the DM physics; 4) the spatially dependent velocity dispersion of the DM due to substructure plays an important role both in considering Sommerfeld or p-wave enhanced annihilation signals, and simulations can provide this information.

The contribution of DM annihilation/decay processes to the Fermi-LAT γ-ray signal in

the Galactic center (or its expected signal from subhalos in the Galactic halo) is difficult to quantify without knowing the Galactic distribution of dark matter and the dark matter particle properties leading to annihilations or decays. To compare the observations with predictions of DM emission, we must rely on numerical simulations that follow the formation of cosmic structure in the highly nonlinear regime. A few high-resolution simulations of Milky Way-sized halos have been completed over the last few years: the Aquarius project,[32] the Via Lactea series,[1] and the GHALO run.[33] The capability of the Fermi-LAT satellite to detect DM annihilation signals from Galactic subhalos has been previously studied by Ref.[43] and Ref.[3] using the high-resolution Via Lactea II (VLII) N-body simulation. To date, there has been no clear signal of dark matter annihilations or decays in the outer Galactic halo. There are numerous possible explanations for nondetection in the outer Galactic halo combined with the observed DGE in the Galactic center, including simulated DM structure being inconsistent with the actual Galactic structure, the predicted DM annihilation/decay luminosity from particle physics being overly optimistic, or some misunderstanding of the observational data. Motivated by the range of possible explanations for the diffuse Galactic gamma ray emission, we present a study of various possible Galactic DM contributions to the observed DGE emission. We construct several maps of gamma-ray emission from different DM annihilation/decay scenarios to illustrate how each scenario may offer additional templates to fit the Fermi data. We additionally compared DM annihilation signals in each scenario to the DGE model from the GALPROP code of[23] in different regions on the sky.

We organize the rest of this chapter as follows: in Section 2.1, we describe different DM annihilation/decay schemes. Our data processing method and γ-ray generating algorithm

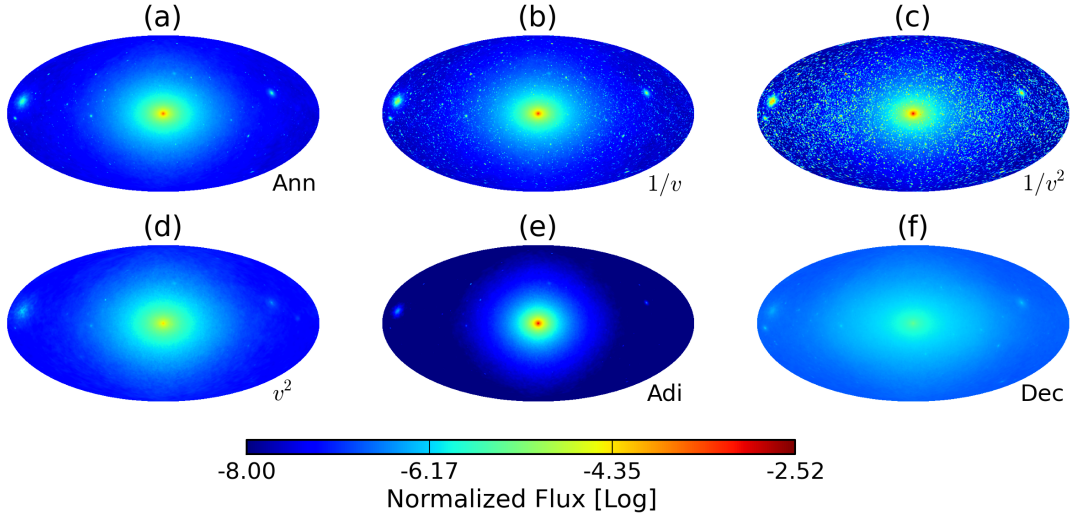

Figure 2.1: Mollweide projection of DM annihilation flux. *From (a) to (f)*: pure annihilation, annihilation with Sommerfeld enhancement $1/v$ correction, annihilation with Sommerfeld enhancement $1/v^2$ correction, annihilation with v^2 correction, annihilation with adiabatic contraction, pure decay. A Gaussian filter with FWHM 0.5° is applied. The flux is normalized such that the host halo has flux unity.

are described in Section 2.2. We show how we add our correction to the normal annihilations/decays in Section 2.3. The all-sky maps results and comparisons to Fermi data for different sky regions are presented in Section 2.4. Conclusions and discussion are presented in Section 2.5.

2.1 Dark Matter Annihilation And Decay

Gamma-rays are one of the final products from DM self-annihilations or decays. In the case of annihilations, the γ-ray flux (photons cm^{-2} sr $^{-1}$s^{-1}) in a solid angle dΩ of a given

line-of-sight (θ, ϕ) can be written as

$$\Psi_\gamma(\theta,\phi) \propto \langle\sigma v\rangle \int_{los} \rho(\mathbf{r})^2 dr, \tag{2.1}$$

where $\langle\sigma v\rangle$ is the velocity-weighted thermal cross section, which is usually treated as a constant, and $\rho(\mathbf{r})$ is the density of DM at position $\mathbf{r} = r(\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$, and r is the distance from the DM particle to the observer. In the case of decay, the γ-ray intensity is then proportional to ρ,

$$\Psi_\gamma(\theta,\phi) \propto \frac{1}{\tau} \int_{los} \rho(\mathbf{r}) dr, \tag{2.2}$$

where τ is the decay lifetime. Eqn-(2.1) and Eqn-(2.2) are usually written as a combination of a particle physics factor P and an astrophysical factor J, while we combined them as Ψ to indicate what has been considered in our γ-ray producing code. Since $\langle\sigma v\rangle$ and τ are unknown, the normalization is arbitrary. For a given astrophysical DM distribution, the Ψ factors given here determine the relative intensity of the predicted flux. Calculating the absolute value of Ψ requires the absolute value of the particle physics parameters. As shown in several papers,[44,45] $\langle\sigma v\rangle$ is less than $\sim 10^{-26}$ cm^{2}s^{-1}, and particle mass M_χ varies from 100 GeV to 10 TeV for the case of WIMPs. For our purpose, only the relative intensity is needed. The γ-ray maps in this work are produced from the simulation data (see Section 2.2) by using one of these formulae in addition to possible cross-section correction factors. We consider in total 6 different γ-ray map production scenarios: (a) pure annihilation without any correction factors, assuming $\langle\sigma v\rangle$ is a constant; (b) annihilation with a Sommerfeld

enhancement of $1/v$ included in the annihilation cross section (v is the relative velocity of the annihilating particles); (c) annihilation with a Sommerfeld enhancement of $1/v^2$, as expected near resonance; (d) annihilation proportional to v^2, consistent with s-wave channel suppression ; (e) adiabatic contraction to the dark matter density profiles, and (f) pure decay, taken to be linear in dark matter density. The correction cases are described in Section 2.3.

2.2 Simulation Data and Map Production Algorithm

The Via Lactea series includes some of the highest resolution collisionless simulations of the assembly of a Milky Way-sized halo. VLI, the first of the series, contains 2×10^8 DM particles, covering the virial volume and surroundings of a host halo of $M_{200} = 1.77 \times 10^{12}\mathrm{M}_\odot$. The host halo and subhalo properties are presented in,[46],[47] and.[48] For this work, we used the second generation of the simulation series, the VLII run,[1,3] which has a slightly higher resolution. It employs about 1 billion particles each of mass $4,100\mathrm{M}_\odot$ to simulate a Milky-Way-sized host halo and its substructures. We extract the roughly 4×10^8 particles of the $z = 0$ snapshot within r_{200} (402 kpc, the radius where the density is 200 times larger than the critical density ρ_c) of the host halo. About 5×10^4 individual subhalos have been identified in this region. The halos found within r_{200} show considerable self-similarity. An generalized NFW profile fit to the density profile,

$$\rho(r) = \frac{\rho_s}{\left(\frac{r}{r_s}\right)^\gamma \left[1 + \left(\frac{r}{r_s}\right)^\alpha\right]^{\frac{\beta-\alpha}{\alpha}}}, \tag{2.3}$$

Ref.[1] gives best-fitting parameters $\gamma = 1.24$, $\alpha = 1$, $\beta = 4 - \gamma$ and $r_s = 28.1$ kpc and $\rho_s = 3.50 \times 10^{-3} \mathrm{M}_{\odot}\mathrm{pc}^{-3}$. The host halo is not spherically symmetric but is ellipsoidally shaped. The detailed host halo properties can be found in Ref.[1]

To produce the γ-ray maps, we follow Ref.[3] and locate a fiducial "observer" 8 kpc from the host halo center along the intermediate principle axis of the ellipsoid. Given the position of an observer, the density field of each particle can be represented by a Dirac-δ function. Eqn-(2.1) and (2.2) can be rewritten as the summation of flux over each particle, i.e. $\sum_i F_i$. For annihilation, $F_i \equiv m_i \rho_i / 4\pi r_i^2$; and decay, $F_i \equiv m_i / 4\pi r_i^2$, where r_i is the distance from the particle to the observer. To avoid shot noise, we smoothed the particles with a SPH kernel so that each particle is represented by a sphere of radius h_i instead of a point. The flux F_i, correspondingly, spreads out as a Gaussian on the sky. We computed ρ_i, h_i and σ_i (the velocity dispersion) by finding the volume encompassing the nearest 32 particles (using the code SMOOTH). The velocity dispersion σ_i is read out simultaneously from this process.

Our final result for the omnidirectional γ-ray signal has been calculated using HEALPIX.[49] We use a NSIDE=512 to model the angular resolution of Fermi-LAT (roughly 0.2°). This is an extremely compute-intensive task, it takes more than 8 hours on a regular CPU to produce a single map with the 400M particles. In order to be able to create the maps involving the different annihilation scenarios and different viewpoints, we have developed a novel GPU-based algorithm.[50]

In this method, we projected $F_i(\theta, \phi)$ onto two separate tangential planes of the two

celestial hemispheres using stereographic projection, then remapped these to the final HEALPIX projections. We have realized that the problem can be reduced to rendering a projected density profile of a particle, weighted by different factors depending on the physics of the annihilation. As these computations became quite similar to those used in visualization, we were able to implement them in the shader language of OpenGL (Open Graphics Library) on a high-end Nvidia GPU for maximum performance. As a result, we are able to render the projected profiles at a rate of more than 10M particles per second and hence build a map in 24 seconds. This thousand-fold speedup was an essential factor, that enabled us to carry out the large number of numerical experiments needed to complete this paper.

Even with the high resolution of VLII, only a portion of the hierarchal structure of DM clumps is resolvable. In principle, the WIMP DM substructure has mass power all the way down to the mass of the DM kinetic or thermal decoupling scales, of order $\sim 10^{-6}\mathrm{M}_{\odot}$. These unresolved structures will boost the final brightness of the γ-ray emission according to the ρ^2 or ρ dependence. Detailed studies of the boost factor $B(M)$ can be found in Ref.[3,43] However, for the purpose of this paper, we choose to avoid these calculations since the boost factor exists for all cases of γ-ray generation scenarios, and we only care about the relative significance. We note that the Sommerfeld correction we use saturates well within our mass resolution range.

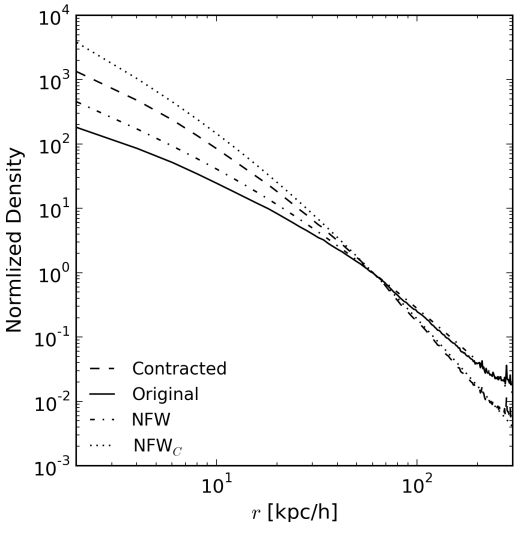

Figure 2.2: The density profile of the host halo and contracted profile. The solid black line is the original density profile of the host halo. After contraction, the profile becomes the long dashed line. The dotted line is the NFW profile with parameters given by Ref.,[1] with $\gamma = 1.24, \alpha = 1$ and $\beta = 1.76$. The dotted dash line is the NFW$_C$ profile describing the adiabatic contraction with $\gamma = 1.37, \alpha = 0.76$ and $\beta = 3.3$. The normalization is chosen such that the intersection point of the contracted and non-contracted profiles are at a density of 1. See text for details.

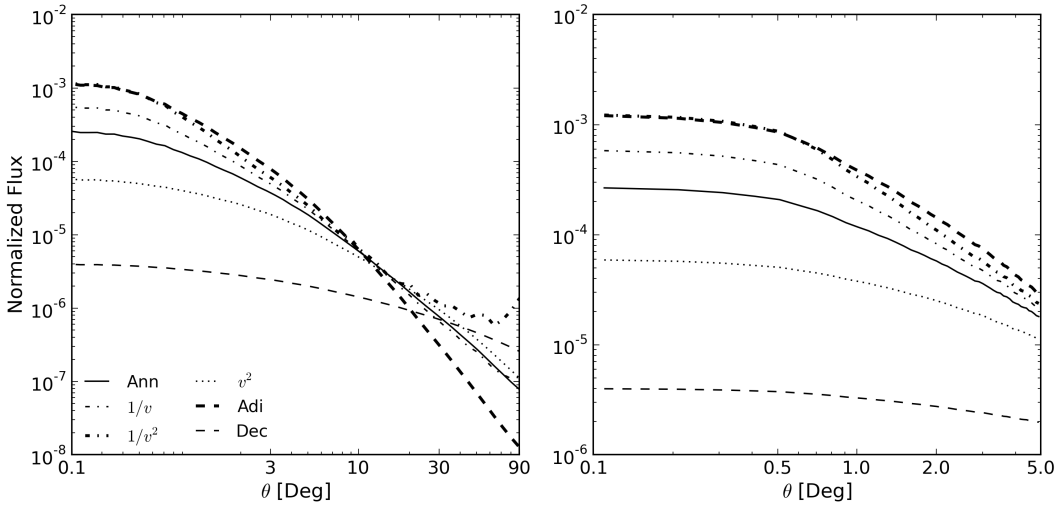

Figure 2.3: The flux annular profile of different cases. *Left panel*: the angular profile of $0 - 90^\circ$ region of the sky maps; *right panel*: the profile of the zoomed in region of the left panel in the inner $0 - 5^\circ$. For legends of lines: solid, pure annihilation; thin dash dotted, annihilation with Sommerfeld enhancement $1/v$ correction; thick dash dotted, annihilation with Sommerfeld enhancement $1/v^2$ correction; dotted, annihilation with v^2 correction; thick dashed, annihilation with adiabatic contraction; thin dashed, pure decay. The flux are normalized such that the host halo has flux unity. See text for details.

2.3 Corrections

2.3.1 Sommerfeld Enhancement

The PAMELA[20] and AMS-02 results[51] show a rise in positron fraction at high energy and a hardening of the spectral index, while no proton excesses were found. If this signal were due to annihilation, the standard thermal cross-section $\langle \sigma v \rangle$ would be too small to simultaneously fit e^+e^- and avoid proton excesses. To gain a larger cross-section, a force carrier ϕ[45,52,53] is proposed for mediating the DM particle interaction. This mediator could be a standard model particle or an unknown boson responsible for dark sector forces. Assuming the DM particle is a Majorana or Dirac fermion, such a process is denoted as $XX \to \phi\phi$. The Sommerfeld enhancement[39] can be therefore calculated. For the resonance case (when $M_X/m_\phi \approx n^2/\alpha$, where α is the coupling strength, and n is an integer), this model gives an enhancement of $S = 1/v^2$ to the cross-section. For the non-resonance case, $S \approx 1/v$. For $c/v \approx 1$, $S \approx 1$. The Sommerfeld cross-section must saturate in a viable model: we take this saturation to be ~ 1 km/s, above the resolution limit of our simulations. From our point of view, this amounts to a renormalisation that we ignore: it is the profile shape and clumpiness that we care about. The cross-section of DM particle pair annihilation changes significantly over the simulation range caused by velocity changes in different environments, i.e. by a factor $\sim 10^2$ or more.[54] In such a case, the subhalos, whose velocity dispersion is less than that of the central halo, light up. For the $1/v^2$ case, the substructures almost dominate, as shown in Fig-2.1(c).

To calculate an accurate Sommerfeld enhancement value at a given location, one needs

the whole phase-space distribution of each particle. This is not possible for our case using the VLII simulation. We therefore follow Ref.[54] and assume a Boltzmann velocity distribution of the relative velocity $v_{\rm rel}$:

$$f(v_{\rm rel}, \sqrt{2/3}\sigma_v) = 4\pi/(2\sqrt{2/3}\pi\sigma_v)^{3/2} v_{\rm rel}^2 \exp[-3v_{\rm rel}^2/\sigma_v^2]. \tag{2.4}$$

The Sommerfeld enhancement is then given by $S(\sigma_v) = \int f(v, \sqrt{2/3}\sigma_v) S(v) dv$. If $S(v)$ has no saturation, then $S(\sigma_v)$ is then $\sim 1/\sigma_v$ for the $1/v$ case and $1/\sigma_v^2$ for the $1/v^2$ case. If one takes into consideration the saturation velocity v_s, below which, $S(v)$ no longer increases, we find $S(\sigma_v)$ presents similar properties. We therefore approximate $S(\sigma_v)$ completely by $S(v)$ without losing much accuracy. In our calculation, we calculated S as $1/\sigma_v$ or $1/\sigma_v^2$ when $\sigma_v > v_s$, and as $S(v_s)$ when $\sigma_v < v_s$. Here σ_v is the velocity dispersion mentioned in the last section. Fig-2.1 (b) and (c) shows the Sommerfeld enhancement case for the γ-ray map. To avoid the cross-section blowing up at very low velocity, we applied saturation velocities equal to 1 km/s for the $1/v$ case and 5 km/s for the $1/v^2$ case.

2.3.2 P-wave Annhiliation

As indicated in the Sommerfeld enhancement case, the relative velocity of particles varies significantly after thermal freeze out. If the cross-section is velocity-dependent, it will either suppress high velocity or low velocity annihilation. Constraints on annihilation processes from CMB and γ-ray observations could be considerably weakened if the s-wave channel were suppressed. P-wave annihilations are orders of magnitudes larger than s-wave

annihilations at recombination and are an interesting case to be considered. For example, neutralino annihilation could be dominated by the p-wave process.[55] In contrast to the Sommerfeld case, we consider an annihilation case whose s-wave process is suppressed. For the annihilation rate, we phenomenologically set $\langle\sigma v\rangle \propto v^{2(n-1)}$, where $n = 1(2)$ for s-wave(p-wave) annihilation.[56] Fig-2.1 (d) shows a map from the p-wave annihilation with the s-wave channel suppressed. More discussion will be given in §2.5.

2.3.3 Baryonic Matter and Adiabatic Contraction

When structure is forming, baryonic matter dissipates its thermal and kinetic energy and falls to the center of the DM halo. The condensation of gas and stars in the inner regions of DM halos will adiabatically contract the DM density distribution and lead to a denser profile in the center. This effect is implied from both theoretical and observational studies.[57] Adiabatic contraction is generally studied using the standard contraction model (SAC) as introduced by Ref.[58] However, hydrodynamic simulations show that the contraction is weaker and a modified adiabatic contraction model was introduced in Ref.[59] We consider the baryonic matter contraction effect only in the host halo, where there is enough baryonic matter to cause the contraction. To simplify the calculation, we follow Ref.[60] and modify the host halo density profile to be a NFW_c profile with $\alpha = 0.76$, $\beta = 3.3$, $\gamma = 1.37$ and $r_s = 18.5$kpc. To conserve the the total mass of the host halo within r_{200}, we use $\rho_s' = 16.83\rho_s$. We convolve a contraction factor $c(r) = \mathrm{NFW}_c(r)/\mathrm{NFW}(r)$ with the density of each particle in the simulation. The mass of each particle is changed accordingly. The contracted density profile and the original profile are shown in Fig-2.2. We normalized the

profile such that the intersection point of the original and the contracted profile is at a density of 1. We include the NFW and NFW$_c$ profile in Fig-2.2 as points of comparison. Although the NFW profile only fits the data well in the outer Galaxy region, the contracted density profile scales the same way as the NFW$_c$ does for the NFW profile. Fig-2.1 (e) shows an annihilation map of the contracted host halo plus uncontracted subhalos.

2.4 All-Sky Maps and Gamma-Ray Signals from Different Sky Regions

The all-sky maps of the six different annihilation/decay scenarios are shown in Fig-2.1. The normalization is chosen such that the host halo has flux unity. All the maps are smoothed by a Gaussian kernel with FWHM 0.5°. They show significant differences in both the center halo profile and the relative subhalo flux. Compared to the pure annihilation case of Fig-2.1a, the $1/v$ and $1/v^2$ Sommerfeld enhancement cases (Fig-2.1b) have considerably more substructures visible. The $1/v^2$ resonant Sommerfeld enhancement case has the largest contribution from substructures to the annihilation flux of the six scenarios. For the $p-$wave annihilation case (Fig-2.1d), the adiabatic contraction case (Fig-2.1e) and pure decay case (Fig-2.1f), substructures are less visible than in the pure annihilation scenario. The smooth component also differs for the six scenarios in terms of both the central maximum and the profile slope, as shown in Fig-2.3. The adiabatic contraction case has a steep central profile such that the normalization of the total flux overwhelms the relatively small contribution from substructure. It looks more spherical in the all-sky map because of the assumption

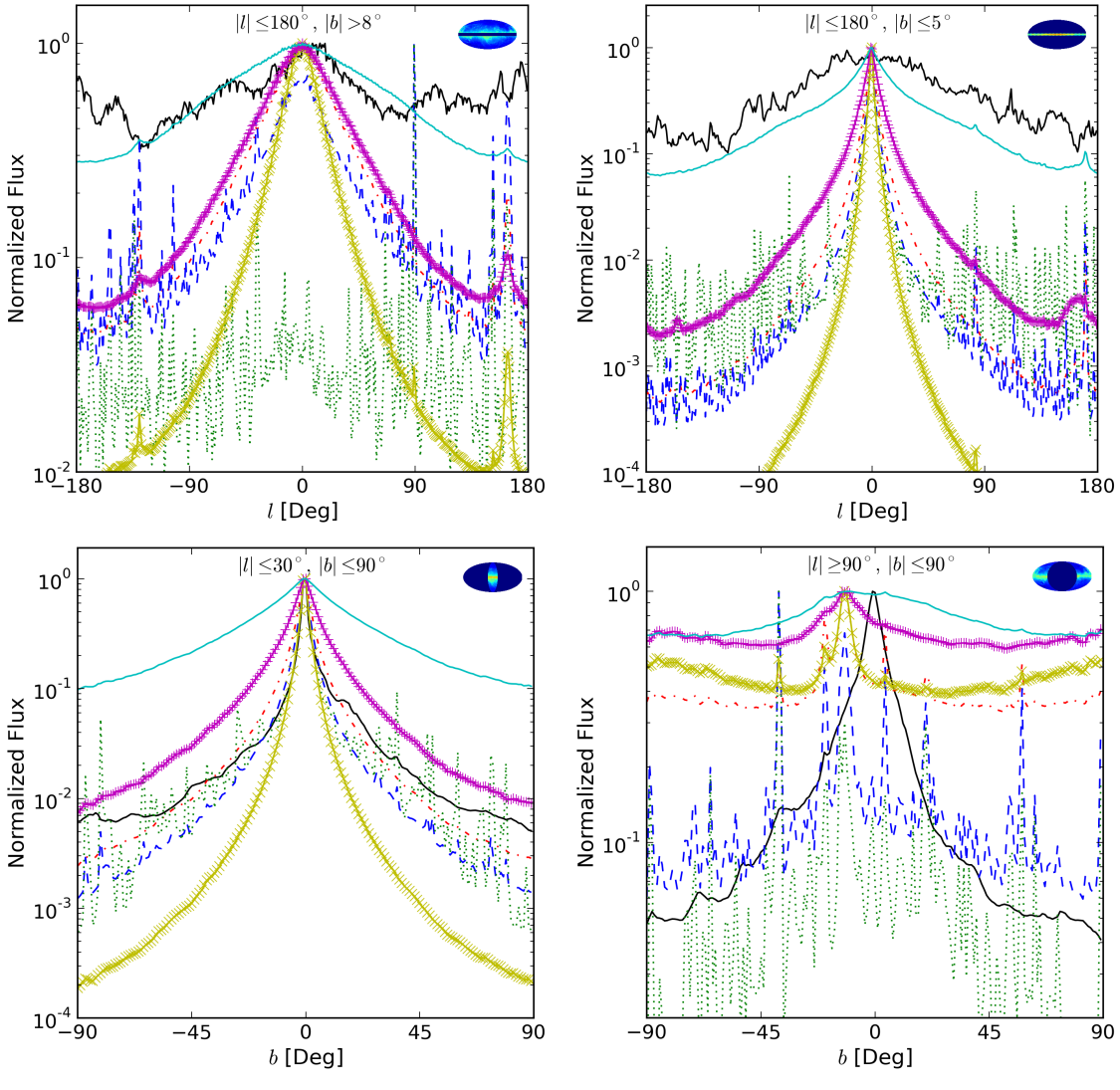

Figure 2.4: Angular profile of different sky regions. *Upper left panel*: the longitudinal profile of $|b| > 8°$; *upper right panel*: the longitudinal profile of $|b| \leq 5°$; *lower left panel*: latitude profile of $|l| \leq 30°$ and *lower right panel*: latitude profile of $|l| \geq 90°$. For legends of lines: black solid, the DGE model ${}^{\rm S}{\rm S}^{\rm Z}4^{\rm R}20^{\rm T}150^{\rm C}5$ from Fermi-data; red dash dotted, pure annihilation; blue dashed, annihilation with $1/v$ correction; green dotted, annihilation with $1/v^2$ correction; solid with "x" symbols, annihilation with adiabatic contraction; magenta solid with "+" symbols, annihilation with v^2 correction; cyan solid, pure decay. All profiles are normalized so that they have maximum unity. See text for details.

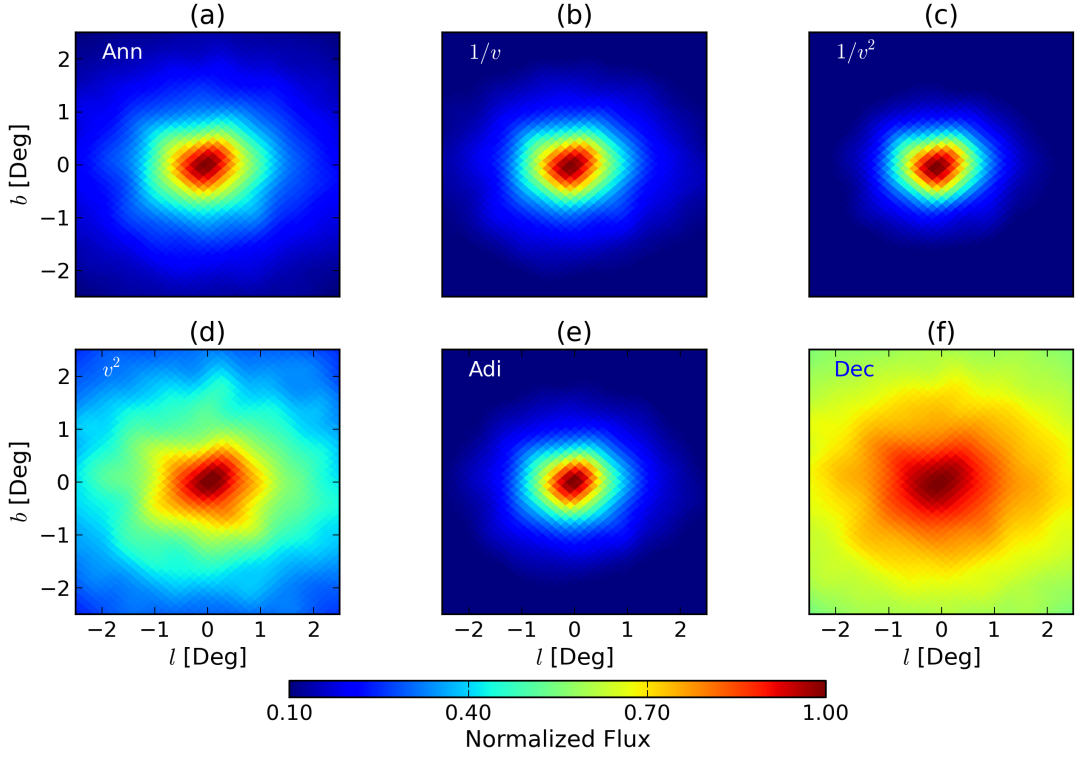

Figure 2.5: Color map of the γ-ray flux in the inner $5^\circ \times 5^\circ$ produced by DM annihilation/decay. From (a) to (f): pure annihilation, annihilation with Sommerfeld enhancement $1/v$ correction, annihilation with Sommerfeld enhancement $1/v^2$ correction, annihilation with v^2 correction, annihilation with adiabatic contraction, pure decay. The flux are normalized such that the center pixel has flux unity.

of spherical NFW$_c$ contraction (Fig-2.1e). Both Sommefeld enhancement cases have a larger central flux than the pure annihilation case due to the distribution of the velocity dispersion. With a v^2 dependence, the $p-$wave annihilation case therefore has a flatter central flux profile. The pure decay case has the shallowest all-sky map and radial profile because of the linear ρ dependence.

We compare our results to the Fermi-LAT DGE model.[23] presented a measurement of the first 21 months of the Fermi-LAT mission and compared a grid of models of DGE emission produced by the GALPROP code. These models incorporate the observed astrophysical distribution of cosmic-ray sources, interstellar gas and radiation fields. They compare the predicted model intensities and spectra with the observations in various sky regions. The models are consistent with the Fermi data in the anti-center regions but underpredict the data in the inner regions of the Galaxy at energies around a few GeV. Their conclusions concerning the discrepancy mainly focused on undetected point sources and cosmic ray spectral index variations throughout the galaxy. Although it is generally accepted that there is an "excess" γ-ray signal most probably from the Galactic Center, it is still worth considering signals from off-center regions.

In,[23] various sky regions are taken into consideration for the comparison of models and observations. We take the same sky regions to plot the angular profiles, namely, $|b| > 8°$, $|b| \leq 5°$, $|l| \leq 30°$ and $|l| \geq 90°$. We also show the DGE model in each subplot. This DGE model is based on the GALPROP model ${}^{\mathrm{S}}\mathrm{S}^{\mathrm{Z}}4^{\mathrm{R}}20^{\mathrm{T}}150^{\mathrm{C}}5$. Here we do not make comparisons to models with other choices of parameters. However, since all the models are required to fit the observational data, they show very similar features. In order to get the

total gamma ray flux we integrated over energy, from 50MeV to 800GeV, in the GALPROP model. The results are shown in Fig-2.4. The upper left panel shows the longitudinal profile of the region of the sky with the central $|b| > 8^\circ$ removed, while the upper right panel shows the longitudinal profile of the central $|b| < 5^\circ$ of the Galaxy. In both panels, the DGE signal shows a relatively flat profile compared to the longitudinal profiles of the five annihilation cases. The longitudinal profile of the decay case provides the closest match to the DGE profile of the six scenarios considered here. The lower left panel shows the latitude profiles of the region with $|l| \leq 30^\circ$. Due to the disk component of the Milky Way the DGE signal has a steep central profile that most closely resembles the latitude profile of the pure annihilation case. In the lower right panel, the latitude profiles of the region with $|l| \leq 90^\circ$ removed shows several interesting features. Each of the annihilation/decay cases has a latitude profile with a off-center peak in contrast with the DGE profile. The off-center maximum for the annihilation/decay cases is due to two apparent large subhalos that are also visible in the all-sky maps (Fig-2.1). The variations in the longitudinal and latitude profiles of the annihilation cases are the result of the contribution from the subhalos to the annihilation signal. The contribution from the subhalos is most important for the two Sommerfeld corrected annihilation cases. The longitudinal profile of $1/v^2$ Sommerfeld corrected case in the upper left panel is completely dominated by emission from subhalos.

We take the inner Galactic region into special consideration since it might be a promising region for detecting DM signals. Other works propose alternative astrophysical models for the γ-ray excess at the Galactic Center, such as unresolved millisecond pulsars.[61][62] strongly disfavor this explanation for both the spatial distribution and spectrum of observed millisec-

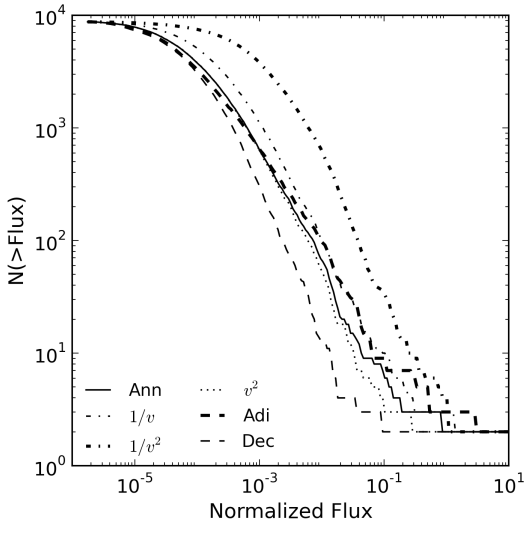

Figure 2.6: The flux function for different cases. *Solid*: pure annihilation. *Thin dash-dotted*: annihilation with Sommerfeld enhancement $1/v$ correction. *Thick dash-dotted*: annihilation with Sommerfeld enhancement $1/v^2$ correction. *Dotted*: annihilation with v^2 correction. *Thick dashed*: annihilation with adiabatic contraction. *Thin dashed*: pure decay. The fluxes are normalized such that the host halo has flux unity. See text for details.

ond pulsars. Assuming that the γ-ray excess mainly comes from DM decays/annihilations, we show here the predicted flux map (Fig-2.5) and angular profile (Fig-2.3 right panel) of the inner $5^\circ \times 5^\circ$ region of the simulated maps for the different γ-ray generating cases. The radial profiles in Fig-2.3 right panel are normalized to the total flux of the host halo, as used in the all-sky maps. The flux maps in Fig-2.5 are normalized such that the center pixel has flux unity and the color scale is linear. The Sommerfeld enhanced (Fig-2.5b and Fig-2.5c) cases and adiabatic contraction case (Fig-2.5e) have a more contracted profile than the pure annihilation case (Fig-2.5a). In contrast, the $p-$wave annihilation (Fig-2.5d) and pure decay case (Fig-2.5c) have flatter profiles illustrated by the greater spread in γ-ray flux over the $5^\circ \times 5^\circ$ region. The six cases of the γ-ray flux maps of Fig-2.5 show considerable variations and complex asymmetrical contours in the central region. This asymmetry would be absent

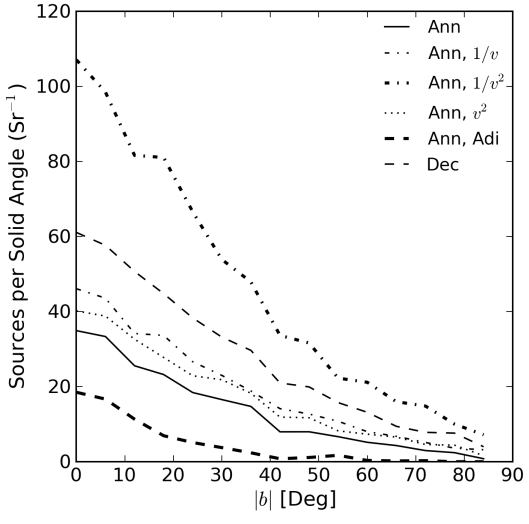

Figure 2.7: The latitude distribution of detected point sources. The detection threshold is chosen such that the maximum number of sources is on the order of 100. Sources are counted in latitude bins of 10°. The legend is the same as Fig. 2.6. See text for details.

from simple parameterized model of the the DM annihilation signal from a NFW/cored profile considered previously.[42]

2.5 Discussions

In addition to the diffuse emission, we are able to use our maps to study the point sources from Fermi-LAT. For annihilations, when we add in the Sommerfeld enhancement, the number density of detected small halos increases. For the $1/v^2$ case, the substructure almost dominates the radiation field other than in the galactic center. Fig-2.6 shows the flux functions of sub-halos for different cases. The flux of the subhalos are calculated by summing over the flux of all pixels inside the halo's angular radius. All the curves converge at $F \to 0$ since the total number of sub-halos in the simulation is fixed at $\sim 10^4$. The curve

for the Sommerfeld enhancement $1/v^2$ is much steeper than the other curves, which indicates that the subhalos are brighter than in other models. We further plot the detectable point sources (with angle extending less than $\sim 2^{\circ}$[63]) in Fig-2.7. Compared to,[63] these curves are shallower than the identified and unidentified sources in the second Fermi-LAT source catalog for the inner region, but are relatively flat at larger latitudes. All of these features imply that observable DM signals could come from off-center subhalos.

In summary, we are not aiming to set new constraints on DM annihilations or decays, but rather to provide possible new templates to fit the DGE data with dark matter models and illustrate the possible range of profiles for plausible models. Compared to a simple assumption that the DM halo has a NFW or an isothermal profile, our results based on the simulation data are more realistic (e.g. Fig-2.2) and include the effects of subhalos. The Galactic central region might be a promising region for detecting a DM signal as recently pointed out by.[64] The various DM maps we present here show significant observational differences when the DM annihilation/decay physics are changed. The morphology of the center region is also asymmetric and complex, and simple analytical templates may fail to match the real signal. There are three key points that we wish to emphasize here.

Firstly, compared to the Fermi DGE result, the annihilation signal, with or without Sommerfeld enhancement, has distinctly different central profiles. The shape of the Fermi excess profile in the Fermi bubble region has recently been used as a means of discriminating dark matter from millisecond pulsar models of the central excess.[65] Our philosophy here is similar but we extend this approach to probing the nature of DM annihilations. We predict different profiles for the alternative dark matter annihilation models. The differences are

especially notable for decaying dark matter, for which the flux profile is very flat.

Secondly, it is clear that some of our models, especially the Sommerfeld cases, provide more bright substructures than are allowed by the observations in terms of unresolved and unidentified point sources. We leave it to others to make detailed comparisons with the data, but our predicted cumulative counts of subhalos should provide the basic template. The $p-$wave annihilation case is a promising model since it removes many of these bright substructures. The model including adiabatic contraction has a similar effect of de-emphasizing the substructure by increasing the contrast between host halo core and the (unadiabatically contracted) subhalos, although the role of adiabatic contraction of the DM in halos remains controversial.

Finally we note that we have not included any spectral signatures, which differ among the various adopted annihilation channels and particle masses: for example the Sommerfeld models require high values for the particle mass ($\sim$ TeV) and hence would be best constrained through observational data at relatively high photons energies.

Chapter 3

Shallower Voids with Warm Dark Matter

This chapter is based on article Yang et al. (2015).[66]

It has been long known[67,68] that the matter of the universe is dominated by DM. There is also overwhelming evidence (e.g.[13]) showing the existence of dark matter (DM), whose nature is still unknown. Although many direct or indirect detection experiments (e.g.[18,69–71]) have occurred and are ongoing, no conclusive result has been reported.

For structure formation, the velocity distribution of DM particles plays an important role. In the standard ΛCDM model, the DM is assumed to be entirely cold from the standpoint of structure formation. The velocity dispersion is negligible at the era of matter-radiation equality ($t_{\rm eq}$), and structure formation proceeds in a bottom-up fashion. Smaller

structures form at first, then larger ones. This model has only a few parameters, that have been determined with high precision. However, several problems remain unsolved on sub-galactic scales. First, the missing satellite problem: simple arguments applied to CDM-only simulations imply that thousands or hundreds of dwarf galaxies are expected in the local group and halo of the Milky Way, however only of order 10 of them were found.[72,73] Second, CDM predicts concentrated density profiles in the central region, e.g. r^{-1} in the NFW[40] profile, whereas many studies of galaxy rotation curves have concluded that the density approaches a constant in the core.[72,74] Third, the number of dwarf galaxies expected in local voids may be less than a CDM model would predict.[75]

Although better modelling of hydrodynamics and feedback processes may solve these problems (e.g.[76]), changing the DM itself could also help resolve some of the issues. Warm dark matter (WDM) has been an attractive alternative since the 1980s (e.g.[77]). Recently, the WDM model has received some interest since it can reproduce all the successful CDM results on large scales, but also solve some small-scale issues. The key feature separating WDM from CDM models is the lack of initial small-scale fluctuations. The WDM has slightly larger velocity dispersion at $t_{\rm eq}$, giving a smoothing of initial fluctuations at a free-streaming length determined by the WDM particle mass. From particle physics, the originally favored WDM candidate was a gravitino (e.g.[78]); more recently, a sterile neutrino[79] has gain attention. Both theoretical and numerical studies (e.g.[80]) have explored WDM models, and observational constraints have been put on the mass of WDM particles. E.g. Ref.[81] give a lower limit of $m_X = 3.3$ keV from Lyman-α forest data (HIRES data). Other independent studies (e.g.[82]) also give consistent limits.

While previous studies on WDM mainly focused on the formation of halos or other dense structures, there has been some work investigating the cosmic web itself. Ref.[83] studied voids in a WDM scenario, but focus on the halo population within them, finding that voids are emptier (of haloes and substructure) with WDM. Below, we study the dark-matter density itself, which follows the opposite trend, growing in density in WDM. Ref.[84] also studied the large-scale-structure traced out by galaxies in a WDM scenario, including a study of galaxy environment. They found that WDM makes very little difference in the usual observables of the galaxy population, when using subhalo abundance matching (SHAM) to identify galaxies. They found that the subhaloes, whose masses are above the mass threshold where halo formation is not substantially disrupted from the loss of power in WDM, are arranged in nearly the same way in WDM and CDM. However, one caveat is of WDM is from hydrodynamic simulations. These simulations indicate that stars may form in filaments instead of haloes if the dark matter is quite warm,[85,86] an issue related to the low "complexity" of dark-matter halo structure in WDM.[87]

Voids are large underdense regions, occupying the majority of the volume of the Universe, and are valuable cosmological probes. For example, via the Alcock-Paczynski test, voids serve as a powerful tool to detect the expansion history of the universe (e.g.[88,89]). Ref.[90,91] have also proposed voids as a probe of modified gravity. The abundance of voids may be sensitive to initial conditions,[92] hence voids may serve as probes of the early universe as well. Previous work has also looked at voids in the context of WDM. Ref.[93] measured the abundance of mini-voids, which become scarcer in a WDM model. Recently, Ref.[94] detected void lensing at a significance of 13σ, raising hopes for void density-profile mea-

surements using lensing. The few-parameter "universal" form for void density profiles that Ref.[95] found will likely help in extracting cosmological information from voids.

In this paper, we study how properties of voids in the cosmic web change in a WDM scenario, with different initial power-spectrum attenuations corresponding to different WDM masses. We analyze these simulations with the ZOBOV void-finder[96] and the ORIGAMI,[97] filament, wall and halo classifier. This chapter is laid out as follows. In section 2, we introduce our warm dark matter N-body simulations. In section 3, we analyze the full cosmic web of DM in a WDM scenario. In section 4, we introduce our void detection methods and show the void statistical properties. In section 5, we show the distinct features of void density profiles for different warmth settings. We give our conclusions and discussion in section 6.

3.1 Simulations

We simulate both CDM and WDM using the GADGET-2[98] code. We use the Zel'dovich approximation[99] to impart initial displacements and velocities at redshift $z = 127$ to particles on the initial lattice of 512^3 particles in a periodic box of size $100\,h^{-1}\,\mathrm{Mpc}$. The initial power spectrum was generated with the CAMB code,[100] using vanilla Λ(C)DM cosmological parameters ($h = 0.7$, $\Omega_\mathrm{M} = 0.3$, $\Omega_\Lambda = 0.7$, $\Omega_\mathrm{b} = 0.045$, $\sigma_8 = 0.83$, $n_s = 0.96$).

To incorporate the effect of a thermally produced relic WDM particle, we apply the

following fitting formula to the transfer functions,[80]

$$T_{\rm WDM} = T_{\rm CDM}(k)[1 + (\alpha k)^2]^{-5.0}, \tag{3.1}$$

where the cutoff scale,

$$\alpha = 0.05 \left(\frac{\Omega_m}{0.4}\right)^{0.15} \left(\frac{h}{0.65}\right)^{1.3} \left(\frac{m_{\rm dm}}{1{\rm keV}}\right)^{-1.15} \left(\frac{1.5}{g_X}\right)^{0.29} . \tag{3.2}$$
$$h^{-1}{\rm Mpc},$$

Here, $m_{\rm dm}$ is the mass of the WDM particle (or the effective sterile neutrino); g_X is the number of degrees of freedom that the WDM particle contributes to the number density (in our case, 3/2). In our set of simulations, we applied $\alpha = 0$, 0.05, 0.1 and 0.2 $h^{-1}\,$Mpc, corresponding first to CDM, and then to WDM particle masses 1.4, 0.8 and 0.4 keV. Note that these masses of sterile neutrinos are disfavoured by Lyman-α forest data,[81] but we adopt them to show the effect of WDM without great computational cost.

Only in the most extreme case of 0.2 $h^{-1}\,$Mpc is α comparable to the interparticle separation. The reason that differences show up even when α is below this scale is the broad shape of the attenuation described in Eq. (3.2). Ref.[80] define a perhaps more meaningful "half-mode" scale radius R_s, via $T(\pi/R_s) = 1/2$; this quantity is $\sim 6\alpha$. We hold σ_8 fixed in the linear power spectrum. This changes the large-scale amplitude, but very slightly since the smoothing kernel acts on scales well below $8\,h^{-1}\,$Mpc.

We emphasize that the differences between CDM and the WDM models would increase

with the mass resolution, because even with "$\alpha = 0$", there is a cutoff in the initial power spectrum from the interparticle spacing of 0.2 $h^{-1}\,$Mpc. To illustrate this, we show in Fig. 3.1 the expected minimum density, as a function of resolution, in a 100 $h^{-1}\,$Mpc box, for CDM and WDM models. We estimate this as follows. As shown below in Fig. 3.3, the spherical-collapse limit[101–103]

$$\delta_{\rm sc} + 1 = [1 - (2/3)\delta_{\rm lin}]^{-3/2} \tag{3.3}$$

accurately gives the transformation from linear to non-linear density on the low-density tail. The lowest linear-theory density in the box will be the expected minimum value of a Gaussian in a sample of $N_{\rm cells} = [(100\,h^{-1}\,{\rm Mpc})/c]^3$ cells, where c is the cell size. This is

$$\delta_{\rm lin}^{\rm min} = -\sigma(c)\sqrt{2}{\rm erf}^{-1}(2/N_{\rm cells}) \tag{3.4}$$

where $\sigma(c)$ is the linear-theory density dispersion in cells of size c. (We approximate $\sigma(c)$ with a spherical top-hat kernel of the same volume as a cubic cell.) As c decreases, the WDM and CDM curves diverge, because $\sigma(c)$ increases in CDM, but not in WDM. The WDM curves do continue to decrease slightly, however, because the number of cells increases.

We have not included the thermal velocity kicks to the individual particles to our simulation, for two reasons. First, particles in the simulation are averages over a statistical ensemble of particles, so it is unclear how to implement the thermal velocity in the initial conditions. Also, this thermal velocity would be negligible for our results. For the α values we use, the RMS velocity distribution is of order 1 km/s,[104] while typical velocities of

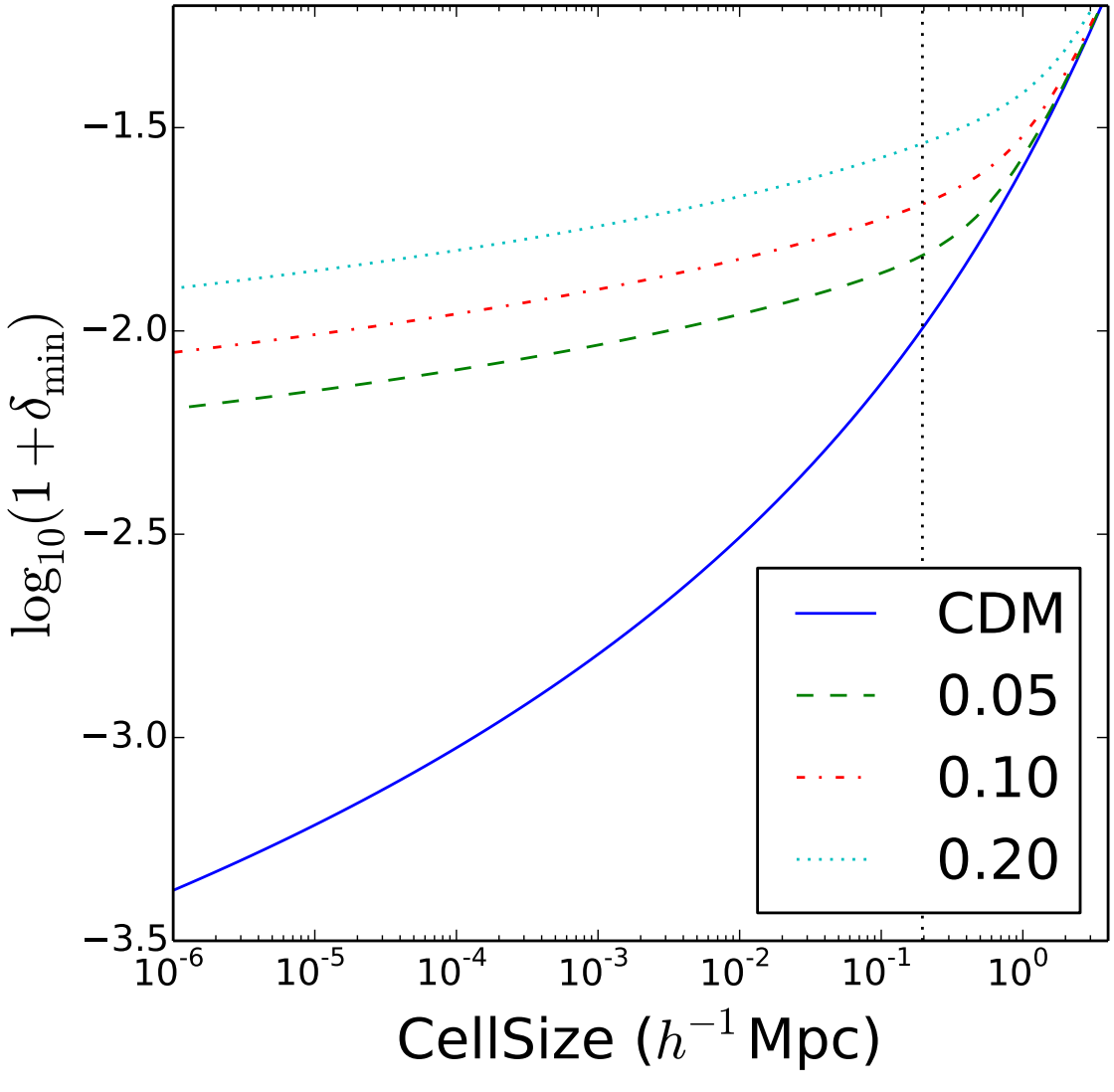

Figure 3.1: The minimum density expected in a $(100\,h^{-1}\,\mathrm{Mpc})^3$ volume, as a function of simulation resolution. The four curves show the four cases investigated below: the WDM cut-off scale $\alpha = 0$ (CDM), 0.05, 0.1, and $0.2\,h^{-1}\,\mathrm{Mpc}$. The cell size is the initial comoving interparticle separation. A dashed line shows the cell size for the resolution used in the simulations used in this work.

particles inside a void are of order 50 km/s at radius around $0.5r_v$.[95] We do not employ a method to prevent spurious fragmentation in filaments,[105] because we focus on voids instead of haloes. Spurious fragmentation should only upscatter particles in morphological type, i.e., turning wall particles into filament or halo particles, and filament particles into halo particles. This artificial fragmentation happens in the presence of anisotropy, but should not happen in voids, where even initially anisotropic volume elements grow nearly isotropic with time.[106]

3.2 The Cosmic Web in a WDM Scenario

Fig. 3.2 shows a slice of an LTFE (Lagrangian Tessellation Field Estimator)[107] density field of the simulation. This estimator makes use of the the fact that, under only gravity, the 3D manifold of DM particles evolve in phase space without tearing, conserve phase space volume and preserve connectivity of nearby points. Hence the sheets (or streams) formed by the initial grids are assumed to remain at constant mass in the final snapshot. Using the initial grid position (or the Lagrangian coordinates) of each particle, the density of each stream could be calculated. We implemented an OpenGL code of LTFE to estimate the density field in a time efficient way. The differences of density fields are clear: small-scale structures are smoothed out in the WDM simulations.

Fig. 3.3 shows mass-weighted 1-point PDFs (probability density functions) at $z = 0$ from the simulations, for each α. We measured the density at each particle using the Voronoi Tessellation Field Estimator (VTFE).[108,109] In the VTFE, each particle occupies a

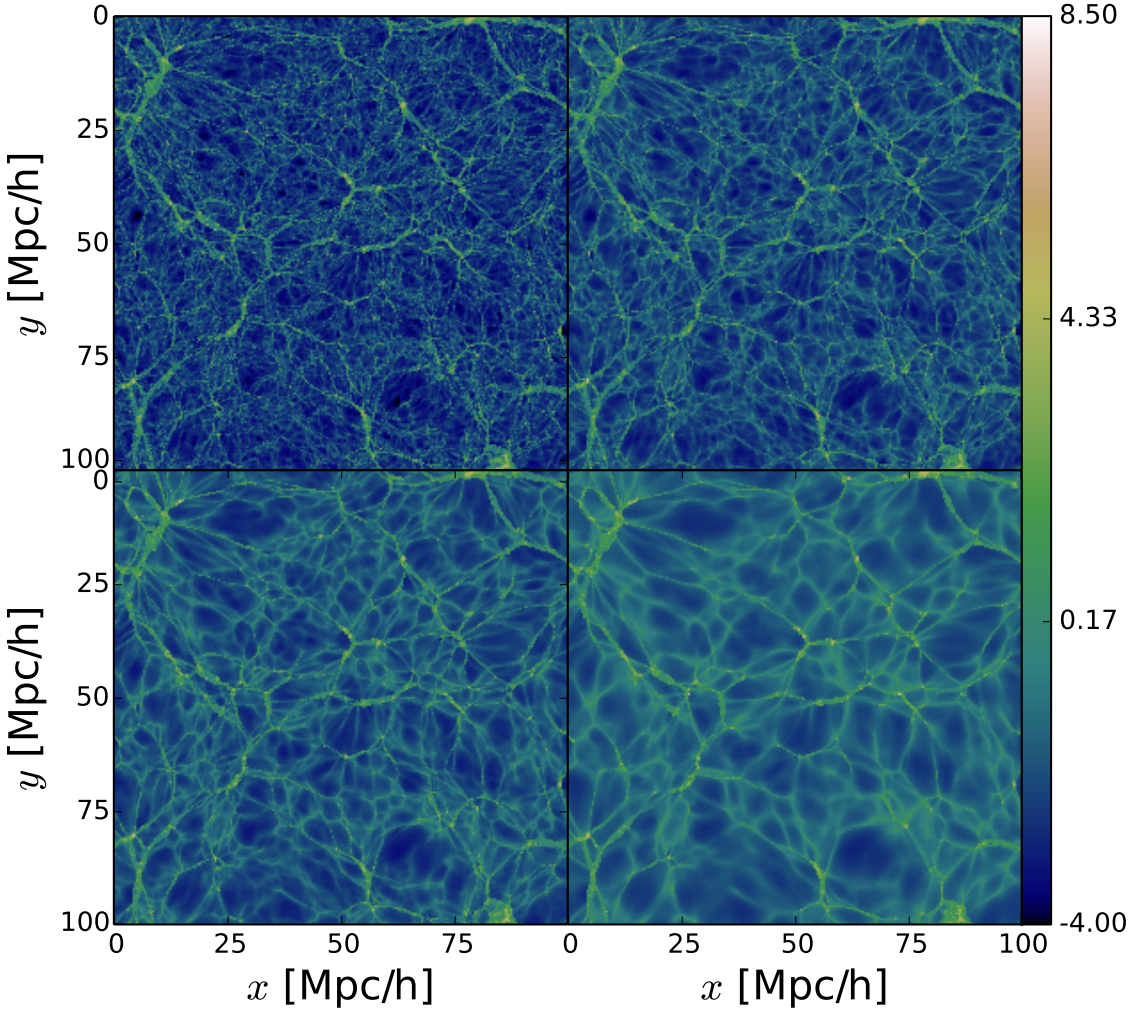

Figure 3.2: LTFE density field slices, showing $\ln(1+\delta)$. From top left to bottom right: CDM, WDM with $\alpha = 0.05\ h^{-1}\,\mathrm{Mpc}$, WDM with $\alpha = 0.1\ h^{-1}\,\mathrm{Mpc}$and WDM with $\alpha = 0.2\,h^{-1}\,\mathrm{Mpc}$.

Voronoi cell, a locus of space closer to that particle than to any other particle. The density $\delta_{\rm VTFE} + 1 = \langle V \rangle / V$ at a particle is set by the volume V of its cell. This density measure is in a sense Lagrangian, but only strictly so without multi-streaming.

At $\alpha = 0$, this mass-weighted PDF shows two clear peaks, noted by Ref.[96] It was already clear that the higher-density peak, a roughly lognormal peak at $\delta \approx e^{6-7} \approx 1000$, comes from halo particles. Ref.[97] firmly established this halo origin by classifying particles with the ORIGAMI algorithm, into void (single-stream), wall, filament, and halo morphologies. This algorithm counts the number of orthogonal axes along which a particle has been crossed by any other particle, comparing the initial and final conditions.

For $\alpha = 0$, the lumpy shape of the total PDF at low densities already suggests that there may be more than 2 components. As α increases, however, an intermediate wall+filament peak becomes unmistakable: the visual impression from the density-field maps that a greater fraction of the matter is in walls and filaments is obvious in the total PDF as well. Again, the ORIGAMI classification confirms this picture. As the WDM mass increases, particles move from low to high densities and morphologies, through the different peaks. A similar effect happens as a function of simulation resolution in CDM: at higher resolution (smaller interparticle separation), the fraction of halo particles increases and fraction of void particles decreases. However, no obvious intermediate wall+filament peak appears in the total CDM PDF.[110] The fractions of particles in walls and filaments remain about constant, with most of the change in particle morphologies appearing in the void and halo peaks . There is a change in the mean log-densities of the wall and filament peaks, however; while a substantial fraction of wall particles have $\delta < 0$ with CDM, note that only the end of the wall tail has

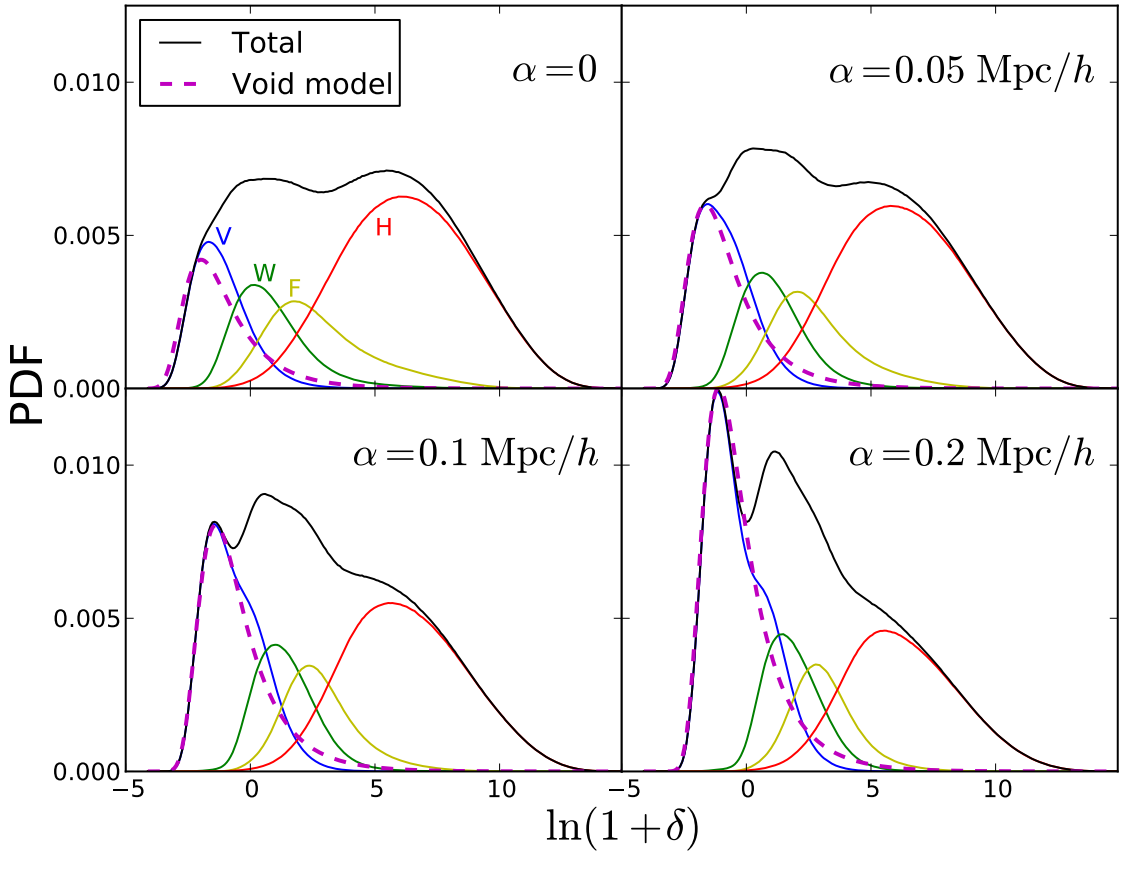

Figure 3.3: PDFs of particle densities for the four simulations. A distinctive peak arises from filament and wall particles at middling densities in a WDM scenario. The V, W, F, and H curves add up to the total: they separate out void, filament, wall, and halo particles, with crossings along 0, 1, 2, and 3 orthogonal axes. $\alpha = 0$ corresponds to CDM; as α increases, the WDM becomes warmer. In this mass-weighted PDF, each particle (Lagrangian element of initial spacing $0.2\,h^{-1}\,$Mpc) enters once. The dashed magenta curve shows the expression in Eq. (3.5).

Figure 3.4: $z = 0$ Voronoi particle densities $\ln(1 + \delta)$ on a 2D Lagrangian sheet, with one particle per $0.2\,h^{-1}\,$Mpc comoving Lagrangian pixel. Each panel shows a 2D, 256^2 slice, a quadrant of a full 512^2 slice.

$\delta < 0$ when $\alpha = 0.2$.

We note that this complicated PDF shape is likely poorly constrained by its first few moments, certainly not the moments of the overdensity δ, and likely not even the log-density $\ln(\delta + 1)$, the x-axis of the plot (e.g.[111,112]). This is a case in which analyzing the instead of the first few moments would be prudent (e.g.[113,114]).

Also plotted is a successful analytic expression for the distribution of void-particle densities,[102,103] derived from a low-$\Omega_{\rm M}$ limit (especially valid for voids) to the evolution of an average mass element.[101] The PDF of the log-density $A \equiv \ln(1 + \delta)$ is

$$P(A) = f_{\rm void} \frac{\exp\left[-\frac{2}{3}A - \frac{(3/2)^2}{2\sigma^2}\left(e^{-(2/3)A} - 1\right)^2\right]}{\sqrt{2\pi\sigma^2}}, \tag{3.5}$$

where σ^2 is the linearly-extrapolated initial variance in cells of size the initial interparticle spacing. $f_{\rm void}$ is the fraction of void, single-stream particles, as measured by ORIGAMI. σ^2 is the variance in spheres of radius $L/N/(4\pi/3)^{1/3}$ (where $L = 100\,h^{-1}\,$Mpc is the box size, and $N = 512$) as calculated from a CAMB linear power spectrum at $z = 0$, truncating the power spectrum to zero at $k > \pi/(L/N)$.

In Figs. 3.4 and 3.5, Lagrangian-space density maps show where these various density regimes appear in the cosmic web. Here, each pixel represents a particle, arranged on its initial lattice. In Fig. 3.5, wart-like blobs within black contours are haloes; these contract

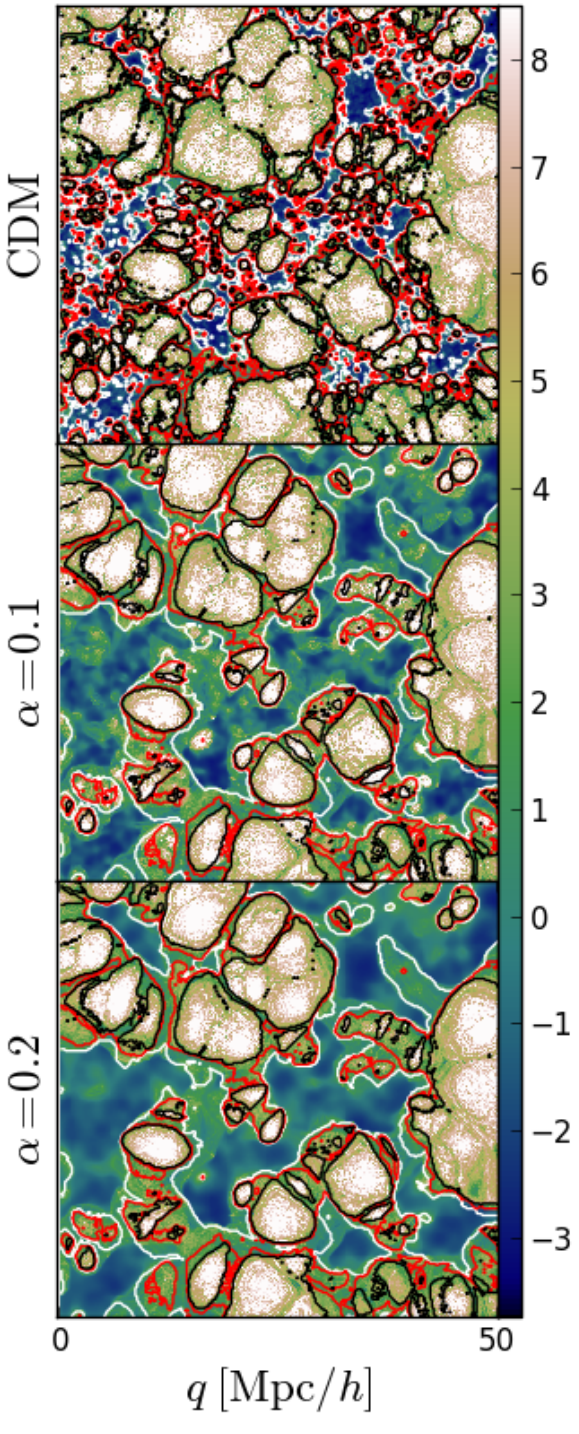

Figure 3.5: Same as Fig. 3.4, with ORIGAMI morphologies added: black, red, and white contours separate void, wall, filament, and halo particle morphologies.

substantially in the mapping to comoving $z = 0$ Eulerian space. The regions within white contours are void regions, which expand in comoving coordinates and come to fill most of the space; see[97] for more detail and an alternative plotting method.

Regarding the topology of the void region, an increase in α decreases the amount of stream-crossing. At high α, even though filaments and walls become more evident visually in the density field, the decreased stream-crossing makes the percolation of the void region even more obvious than in CDM.[110] Even in the 2D Lagrangian slice in Fig. 3.5, the void region obviously does pinch off into idealized convex voids (e.g.[115,116]).

3.3 Void Detection and Properties

ZOBOV first uses a Voronoi tessellation to get the density of each particle. After that, it uses each local density minimum as a seed and groups other particles around it using the watershed algorithm, forming a "zone", regions with a density and a ridge. These zones are combined into larger parent voids using essentially another application of the watershed algorithm, giving a hierarchy of subvoids.

In the default "parameter-free" algorithm output, the whole field is a single large super-void with many levels of sub-voids, which is difficult to use directly. We therefore required the zones added to a void to have core density less than $\bar{\rho}$, the mean density of the simulation. ZOBOV measures the statistical significance of voids, compared to a Poisson process. The probability a void is real depends on the density contrast, defined as the ratio of the minimum density on the ridge separating the void from another void to the void's minimum

density. To focus on voids with low discreteness effects, we analyze voids with significance larger than 3σ according to this density contrast criterion and measure their properties. We found about 7600, 6700, 3900, and 1400 voids for the 4 simulations. Consistently with the results of Ref.,[93] the number of voids decreases as the cut-off scale increases. Compared to,[95] our much higher mass resolution (compared to their sparse-sampling to better approximate a galaxy sample) allows much smaller voids to be detected in the matter field. We acknowledge that these smallest structures would likely not show up in a galaxy survey unless it was extremely deep. Even then, small voids may be smeared out by redshift-space distortions. However, some of these small (sub)voids would be in low-density regions, with smaller redshift-space distortions smearing them out. And also, we emphasize that many of these small void cores, at low sampling, would also likely be centers of larger, parent voids, that are not necessarily included in our pruned catalog.

Comparing to Ref.,[117,118] which considered the hierarchical structure of voids in different levels of smoothing, we do not explicitly take out the sub-voids in each level of the hierarchy but use their effective sizes to characterize them. Apparently, larger voids show up in lower levels of the hierarchy tree. Fig. 3.6 shows distributions of void effective radius, $r_{\rm eff}$, defined via $V_r = 4\pi/3 r_{\rm eff}^3$. The size of the voids in our simulations peaks roughly at 2 h^{-1} Mpc, and shifts slightly outward at high α (moving to WDM). The abundances of voids around the interparticle spacing, $0.2\,h^{-1}$ Mpc, are very sensitive to discreteness noise. Thus, we truncate each plot on the left at twice the mean particle separation, at $\sim 0.4\,h^{-1}$ Mpc.

In the right tails ($r_{\rm eff} \gtrsim 3$ Mpc) of Fig. 3.6, there are more large voids in a WDM scenario, but not dramatically so. This may depend more on the void definition than

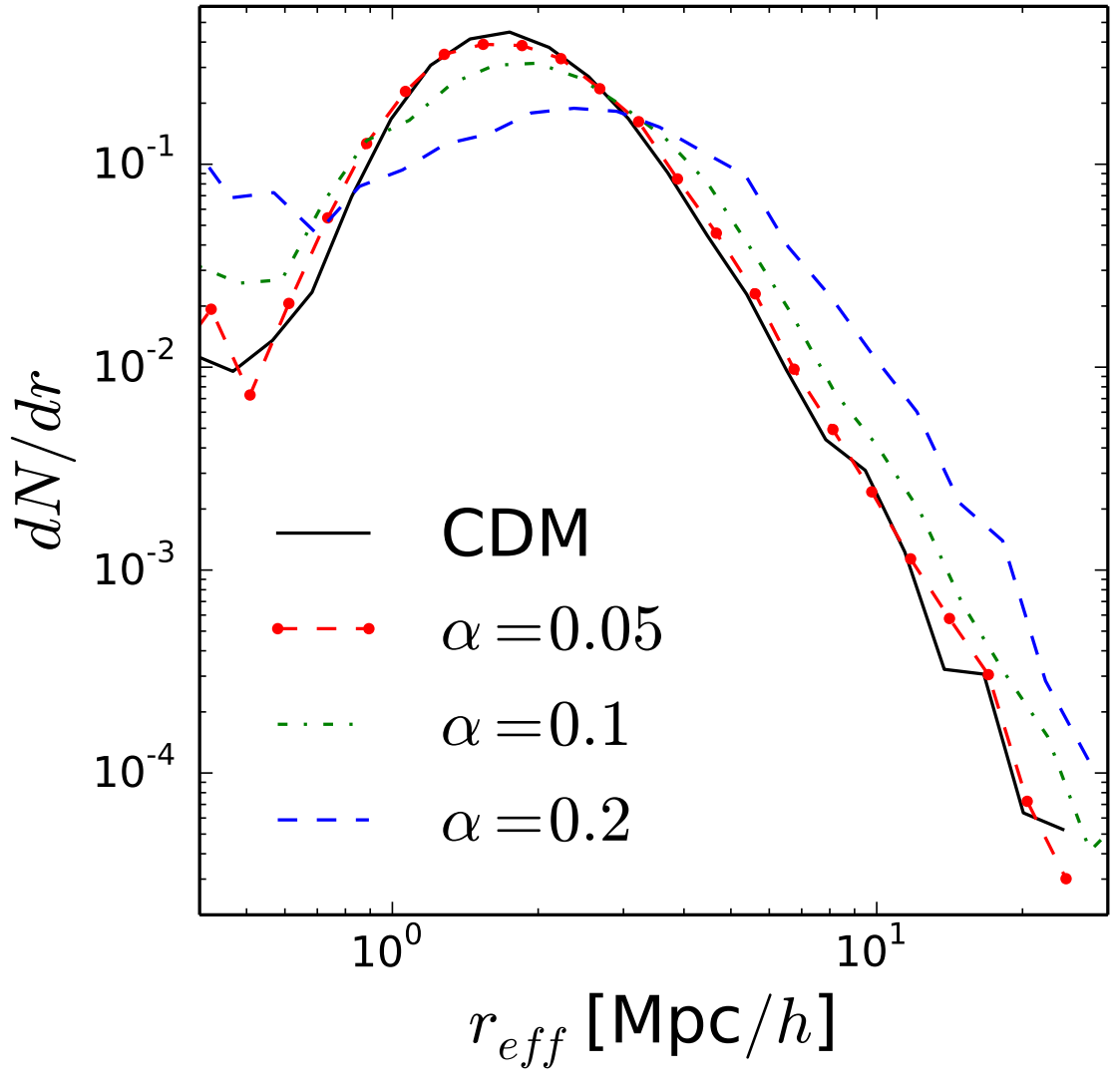

Figure 3.6: The void radius $r_{\rm eff}$ distribution for different dark matter models.

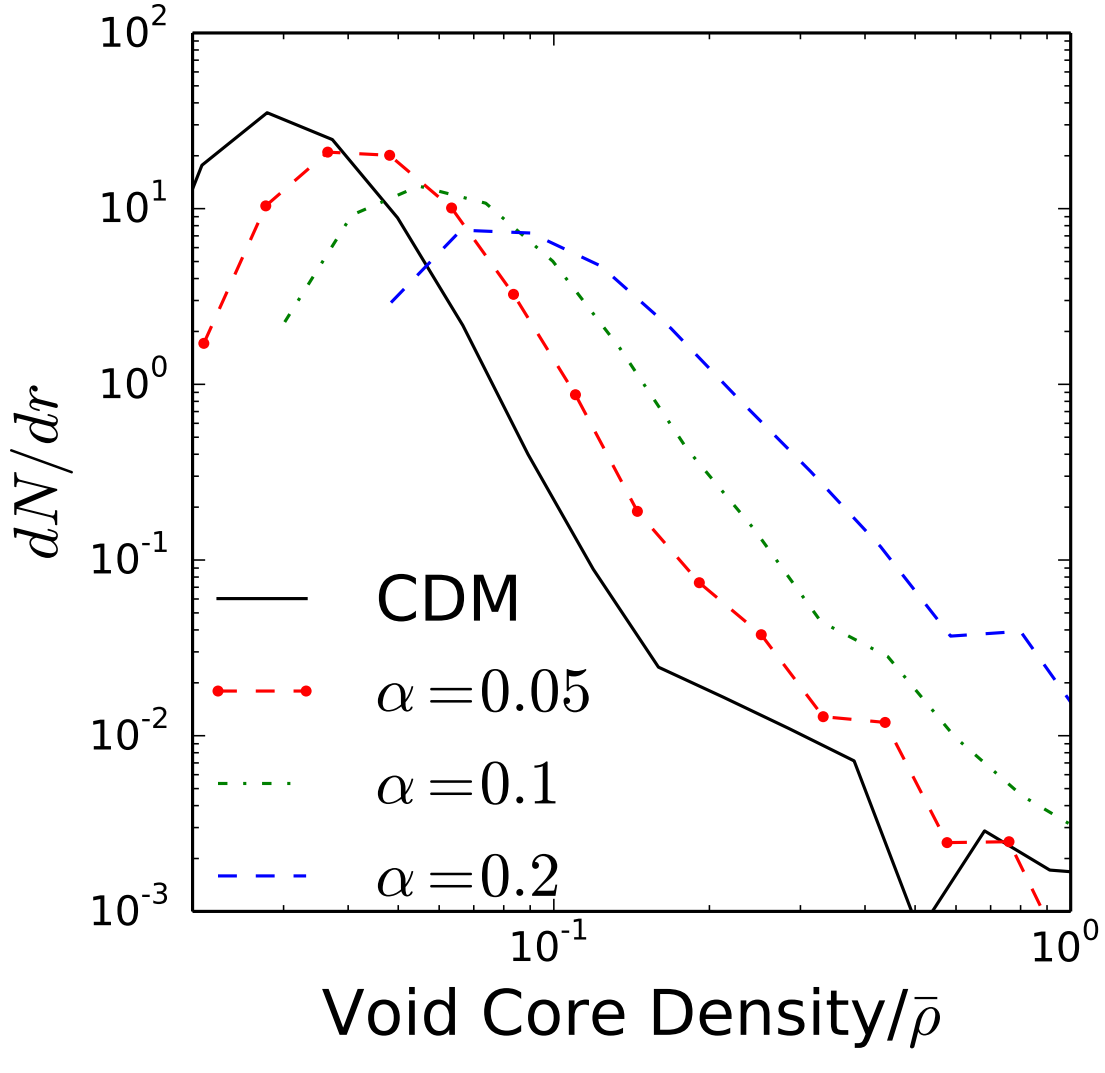

Figure 3.7: The void core density distribution for different dark matter models.

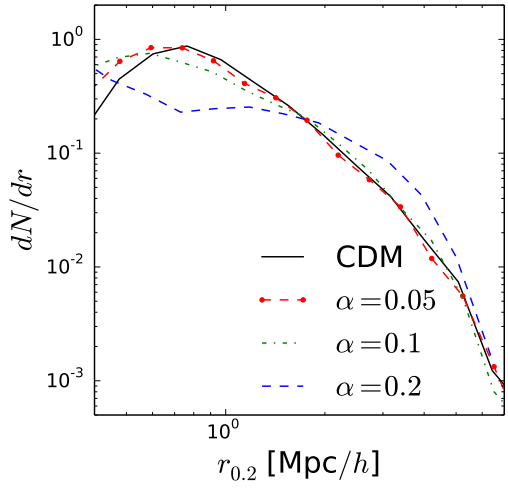

Figure 3.8: The distribution of $r_{0.2}$ for different dark matter models.

the distribution of core densities, however, since the reported radius of a void obviously depends crucially on where its boundary is drawn. While at radius $\gtrsim$ 1Mpc, the abundance depression on the left and increasing at the right of the figure clearly shows the effect that sub-structures in larger voids have been suppressed. Fig. 3.7 shows the minimum "core" Voronoi density distribution for voids and sub-voids (in units of the mean density $\bar{\rho}$). It shows that the centers of voids become shallower in WDM. This is a simple physical effect: in WDM, the initial density PDF on the scale of the interparticle spacing is narrower than in CDM, because of the small-scale attenuation. This results in a narrower particle density distribution at $z = 0$, as shown in Sec. 3. In particular, density minima, in the low-density tail, increase in density.

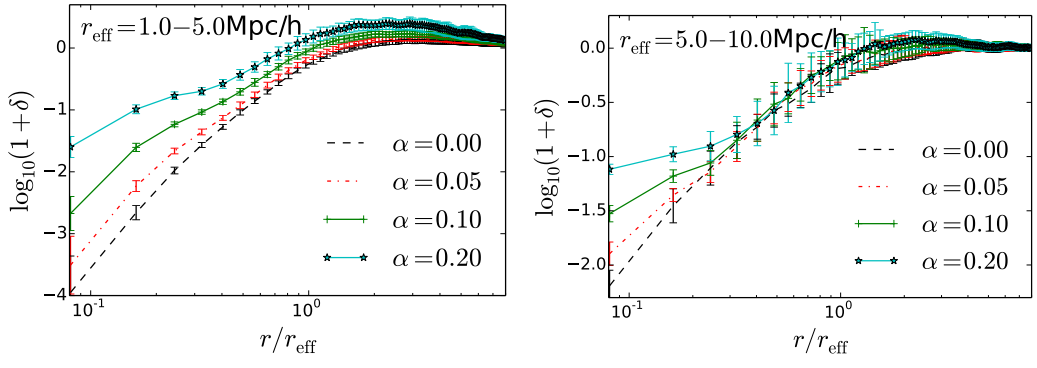

Figure 3.9: Void density profiles measured and scaled with $r_{\rm eff}$. The origin is the density minimum of each void. The left panel shows voids in the 1-5 h^{-1} Mpc radius bin; the right panel shows results from 5-10 h^{-1} Mpc voids. Error bars show the 2σ error, dividing by $\sqrt{N}$, where N is the number of stacked voids.

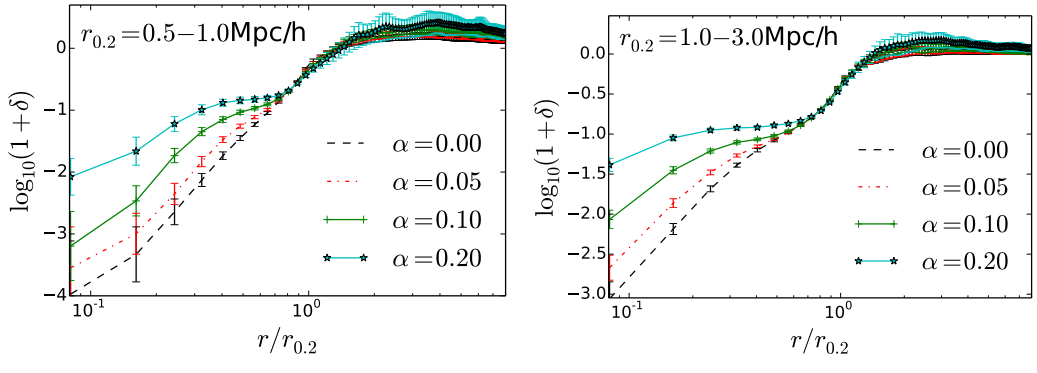

Figure 3.10: Void density profiles as in Fig. 3.9, except scaled with $r_{0.2}$. The left panel shows the voids in the $0.5 - 1.0\,h^{-1}$ Mpc radius bin; the right panel shows the $1.0 - 3.0\,h^{-1}$ Mpc bin.

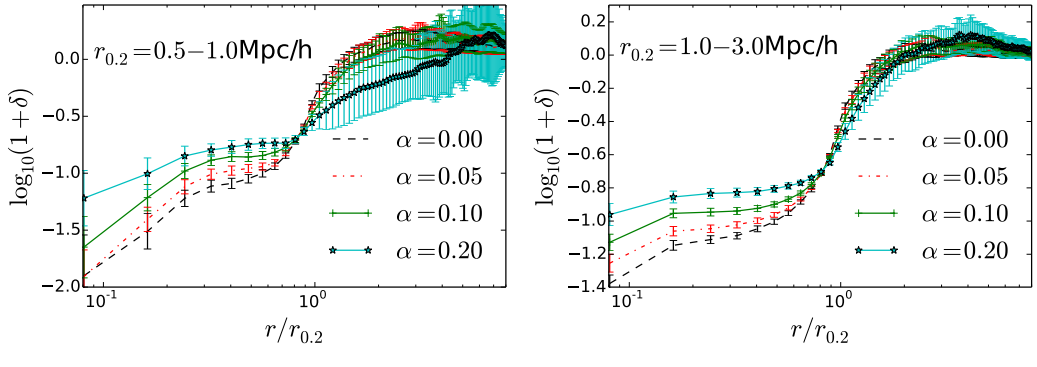

Figure 3.11: Void density profiles as in Fig. 3.10, except using void volume centroids as centers.

3.4 Density profiles

We take special interest in void density profiles, as they have been measured by several different authors recently.[94,95,119–121] We show density profiles in Fig. 3.9. As found by Ref.,[95] smaller voids are, on average, deeper.

Our voids are mostly in the radius range 1-10 $h^{-1}\,$Mpc. We divide the voids into two bins of effective radius, 1-5 and 5-10 $h^{-1}\,$Mpc. We measure the density profiles starting from void centers using linear radial bins. For radial bin $[r, r+\Delta r)$, the density is simply $3N_r/4\pi[(r+\Delta r)^3 - r^3]$ where N_r is the number of particles detected in this bin. We investigate two definitions of the center: a) the actual density minimum of the void, as measured by the VTFE; and b) the volume centroid of the void, defined as $\sum \mathbf{x}_i V_i / \sum V_i$, where $\mathbf{x}_i$ is the position of particle i belonging to the void, and V_i is the Voronoi volume of that particle. The volume centroid would be easier to locate observationally than the density minimum. These profiles are further scaled by different radii r_s (i.e. $r_{\rm eff}$ or $r_{0.2}$) using linear interpolation. Note that for the profiles starting from the density minimum, the Voronoi tessellation always guarantees a particle in the center, and hence a spike. We remove the central bin to remove such an artificial spike. Error bars of each data bin were measured using the standard deviation divided by $\sqrt{N}$, where N is the number of profiles stacked. The error bars shown in this paper are all 2-σ errors.

Since the voids detected by ZOBOV are highly irregular in shape, ZOBOV's effective radius may not be the most meaningful radius measure in all cases. We visually check the voids and found that most of the irregularities are in the noisy edges of the voids. We therefore

define a radius $r_{0.2}$ for voids such that at this point the average density encompassed is $0.2\bar{\rho}$, following Ref.[122] The distribution of $r_{0.2}$ is shown in Fig. 3.8. $r_{0.2}$ was typically from $0.5 - 3\,h^{-1}\,\mathrm{Mpc}$. The distribution of voids of $r_{0.2} \gtrsim 0.5\ h^{-1}\,\mathrm{Mpc}$ shows similar features as does r_{eff} due to the small scale suppression effect of WDM. Note that the number of small voids, with $r_{0.2} \lesssim 0.8\,h^{-1}\,\mathrm{Mpc}$, is curiously higher in WDM than in CDM. For these poorly-resolved small voids, suppose WDM voids are simply shallower versions of CDM voids. The WDM voids will tend to have smaller $r_{0.2}$, since, starting from a higher density minimum, a sphere needs not go out as far to reach an enclosed density of $0.2\bar{\rho}$. Again we use the core particle as the center, scaled the void profiles measured by shell bins by $r_{0.2}$, and stacked them, shown in Fig. 3.10. The central densities are quite similar across different voids. Note that $r_{0.2}$ is typically 2-3 times smaller than r_{eff}. This is surprisingly large, but may be the effect of including the full, generally irregularly shaped, density ridges around voids. As discussed in Ref.,[122] choosing a different definition of void size simply moves voids among radius bins, but does not change their total abundance.

The shallowing of the density profile in central bins is entirely unsurprising if the profiles are measured from density minima; this follows almost trivially from the increase in density minima. Less obvious is the behaviour of void profiles as measured from their volume centroids. The profile from the volume centroid is more observationally relevant, because there is some hope of inferring it from a dense galaxy sample, while locating a 3D density minimum would be quite difficult. In Fig. 3.11, we show density profiles measured in the same way as before, except from the volume centroids using the Voronoi volumes of all particles reported in the ZOBOV void. These profiles are scaled by $r_{0.2}$ and stacked in the

same radius bins as in the previous case. As previously, we tried to use $r_{\rm eff}$ to scale and stack, but in highly noisy profiles. The reason is the same: the shapes of the voids become irregular when their sizes are small, and there is some randomness in whether part of a density ridge is included or not, so the ZOBOV effective radius can be noisy.

Either scaling the profile with $r_{0.2}$ or $r_{\rm eff}$, the density profiles show some universalities – the central part of the profile is relatively stable. For different DM settings, the center part of the profiles is clearly different. While the profiles when scaling by $r_{0.2}$ are less noisy than those scaled by $r_{\rm eff}$, the profiles' shapes are similar in both cases. It is reassuring that the results hold whether the volume centroid or the density minimum is used to measure the density profile.

In most previous profile studies, people use the effective radius $r_{\rm eff}$ to scale the density profile and get universal profiles – a relatively flat central plateau, a sharp edge and a compensated wall, tending to unity faraway. We argue that using $r_{0.2}$, the radius at which the mean enclosed density reaches 0.2, as a scaling constant to find small void profiles is better. The void wall radius deviates from $r_{\rm eff}$ significantly with non-spherical shapes, while $r_{0.2}$ characterizes every void in the context of a spherical-evolution model. If the voids are self-similar, as stated in,[120] $r_{0.2}$ surely returns more consistent profiles. This is indeed shown in Fig. 3.10 and Fig. 3.11. Since the slope of profiles is apparently the largest at $r \approx r_{0.2}$, a lensing measurement would be most sensitive if using this definition.

3.5 Conclusion and Discussion

We measure statistics and density profiles of voids in different dark matter settings, namely CDM, and WDM with characteristic scales $\alpha = 0.05$, 0.1 and 0.2 h^{-1} Mpc. In summary, the voids in WDM are shallower. WDM voids also tend to be larger, although this effect depends somewhat on the void definition. The number of statistically significant voids is also smaller in WDM simulations. The main question this paper poses is whether void density profiles can be used to detect WDM, or some other process that attenuates initial small-scale power. Our answer is yes, in principle.

One advantage of using voids rather than high-density regions is that the structure of voids is much less sensitive to baryonic physics than in high-density regions. This is because matter in voids undergoes no stream-crossing on cosmological scales (e.g.[110]). Substantial differences between dark-matter and baryonic physics are only expected when streams collide; in collisionless dark matter, the streams pass through each other, while streams of gas collide, e.g. forming shocks.

However, it is quite difficult to constrain matter density profiles observationally, even if the voids themselves can be located using galaxies or other tracers, which becomes tricky at small radius because of issues such as redshift-space distortions. Lensing could be used to constrain density profiles, as shown in.[94] However, lensing is only sensitive to the gradient of the surface density, which becomes zero at the very centre of the void. Fortunately, our results do show a difference in the density gradient in different WDM scenarios. Another possible probe is the integrated Sachs-Wolfe effect (e.g.[123]), usefully sensitive to the poten-

tial, although that measure is difficult to detect for small voids because of the dominant primordial CMB. But there are other probes of the potential through voids, such as in fluctuations in the cosmic expansion rate as measured with supernovae. Another aspect of voids that we did not measure, but would also be sensitive to matter density profiles, is the velocity field within voids; perhaps very faint tracers such as absorption lines can be used to constrain these (e.g.[124]). Another possible direction to constrain WDM is through the properties of filaments and perhaps walls in WDM, which we quantitatively showed become more prominent; however, we leave a detailed study of their properties such as density profiles to future work.

Chapter 4

Response of Power Spectra to Initial Spikes in CDM Regime

This chapter is based on Neyrinck & Yang (2013).[125]

Fluctuations in the present-epoch cosmic density field are rich in cosmological information. However, even for the theoretically straightforward real-space dark-matter field, the relation between the initial and final fluctuations is obscure on small scales. In the usual overdensity $\delta = \rho/\bar{\rho} - 1$, the shape of the power spectrum departs substantially from linear theory on small scales. A further obstacle to inferring initial information is the substantial covariance which arises in the δ power spectrum on small scales.[126] This covariance greatly reduces the Fisher information, i.e. increases error bars on cosmological parameters observable in principle.[127–130]

Fortunately, much of this apparently lost Fisher information can recovered using a local 1-point probability density function (PDF) Gaussianizing transform, such as a logarithm or rank-order-Gaussianization.[131–134] These Fisher analyses used measurements of final-conditions covariances in the power spectrum, which are related to the independent degrees of freedom resident there, but not directly to the initial degrees of freedom.

In this Letter, we explicitly track these initial degrees of freedom, in an N-body experiment in which we observe the results of 'ringing' the initial density field with initial power-spectrum spikes at different wavenumbers. This quantifies the 'memory' of initial conditions (ICs) in the final field, but differently than the propagator,[135] which can be thought of as a cross-correlation of a mode's amplitude and phase in the initial and final conditions. Information in an initial mode gets deposited in larger or smaller-scale modes, but does not disappear, as a naive interpretation of a vanishing propagator might suggest.

We measure the response in the power spectra of a few transformed fields: $e^{-(1+\delta)}$; the log-density $\ln(1+\delta)$; the rank-order-Gaussianized density Gauss(δ); and the reciprocal-density $\frac{1}{1+\delta}$. The 'Gauss' function is an increasing function, depending on the PDF of its argument, whose result has a Gaussian PDF. Explicitly, $\text{Gauss}(\delta) = \sqrt{2}\sigma \text{erf}^{-1}(2f_{<\delta} - 1 + 1/N)$, where $f_{<\delta}$ is the fraction of cells less-dense than δ, σ is the standard deviation of the Gaussian that δ's PDF is mapped onto, and N is the number of cells. Below, we set $\sigma = \sqrt{\text{Var}[\ln(1+\delta)]}$. So Gauss($\delta$) is essentially $\ln(1+\delta)$ with its skewness removed.

These transformations increasingly emphasize low-density regions, where the initial fluctuations are the most pristine. We expect the different effective weightings provided by

the transformations to affect the way power migrates in scale as structure forms. Ref.[136] showed analytically using the Zel'dovich approximation[99] that the baryon acoustic (BAO) peak location in the correlation function changes if a transform is applied to the field, based on the density regimes that the transform emphasizes. In δ, $\ln(1+\delta)$, and $\frac{1}{1+\delta}$, the BAO peak location is slightly biased inward, nearly unbiased, and biased outward compared to linear theory. The present Letter carries these results to general power-spectrum features.

We investigate these particular transformations for various reasons. In a lognormal,[137] or nearly lognormal field, two-point statistics of the log-density capture a much greater fraction of the Fisher information in the field than those of δ.[111,112,131] Forcing the one-point moments to zero, as Gaussianization does, generally decreases higher-order multi-point correlations as well, and allows two-point statistics to more effectively capture the information in a field. Also, the power spectrum of the log-density and Gaussianized density have an intriguingly nearly-linear shape, which this study addresses. $\frac{1}{1+\delta}$ is of some theoretical interest, being a simple quantity in a Lagrangian framework (just the Jacobian of the deformation tensor); also, it is rather insensitive to multistreaming, which only further suppresses high-density regions. We investigate $e^{-(1+\delta)}$ because of its relevance to Ly-α-forest measurements; from a dark-matter simulation, it gives the flux, in the (poor) approximation that neutral hydrogen follows dark matter. It is of interest in the study of density transforms, since Gaussianization[138] was tried in Ly-α-forest studies,[139] but later works skipped that step (e.g.[140]). Without noise, $\mathrm{Gauss}(\delta) = -\mathrm{Gauss}(e^{-(1+\delta)})$, so comparing the results of these two transformations is a test of Gaussianization's usefulness in Ly-α studies.

In Section 4.1, we describe the simulation suites and other techniques. In Section 4.2,

we give our main results. In Section 4.3, we explain a crude toy model that qualitatively captures the migration of power in Gaussianized-density power spectra. In Section 4.4, we discuss estimating practical power-spectrum covariance matrices and Fisher information from these results.

4.1 Methods

We ran three ensembles of 24 simulations: one 'control' simulation without initial power-spectrum spikes, and 23 simulations with spikes, each in a different power-spectrum bin. Previous numerical experiments have been run in which some Fourier modes are kept fixed, and others are changed (e.g.[141–144]). Our simulations, run using GADGET2,[98] each have 256^3 particles in a $256\,h^{-1}\,\mathrm{Mpc}$ box, and ICs generated using the ZA at redshift $z = 127$. The simulations used vanilla ΛCDM cosmological parameters: $(h, \Omega_b, \Omega_{\mathrm{cdm}}, \Omega_\Lambda, \sigma_8, n_s) = (0.73, 0.045, 0.205, 0.75, 0.8, 1)$. Each spike was inserted by multiplying the mode amplitudes in the corresponding bin by $\sqrt{2}$, holding all other aspects of the ICs fixed. As usual, each initial mode's phase is random and uncorrelated to other phases. However, to avoid possible effects from spikes that arise randomly from cosmic variance, the amplitude of each initial mode at k is set to exactly $\sqrt{P(k)}$. So, the only difference between the three ensembles is in their sets of random phases. Because Fourier amplitudes are not drawn from a Rayleigh distribution as usual, some higher-order statistics within each box may be suppressed, but the simulations are suitable for their intended purpose, to track power migration.

On the scales investigated, $P_\delta \gg 1/n$, the P_δ shot noise. For the δ analysis, we use

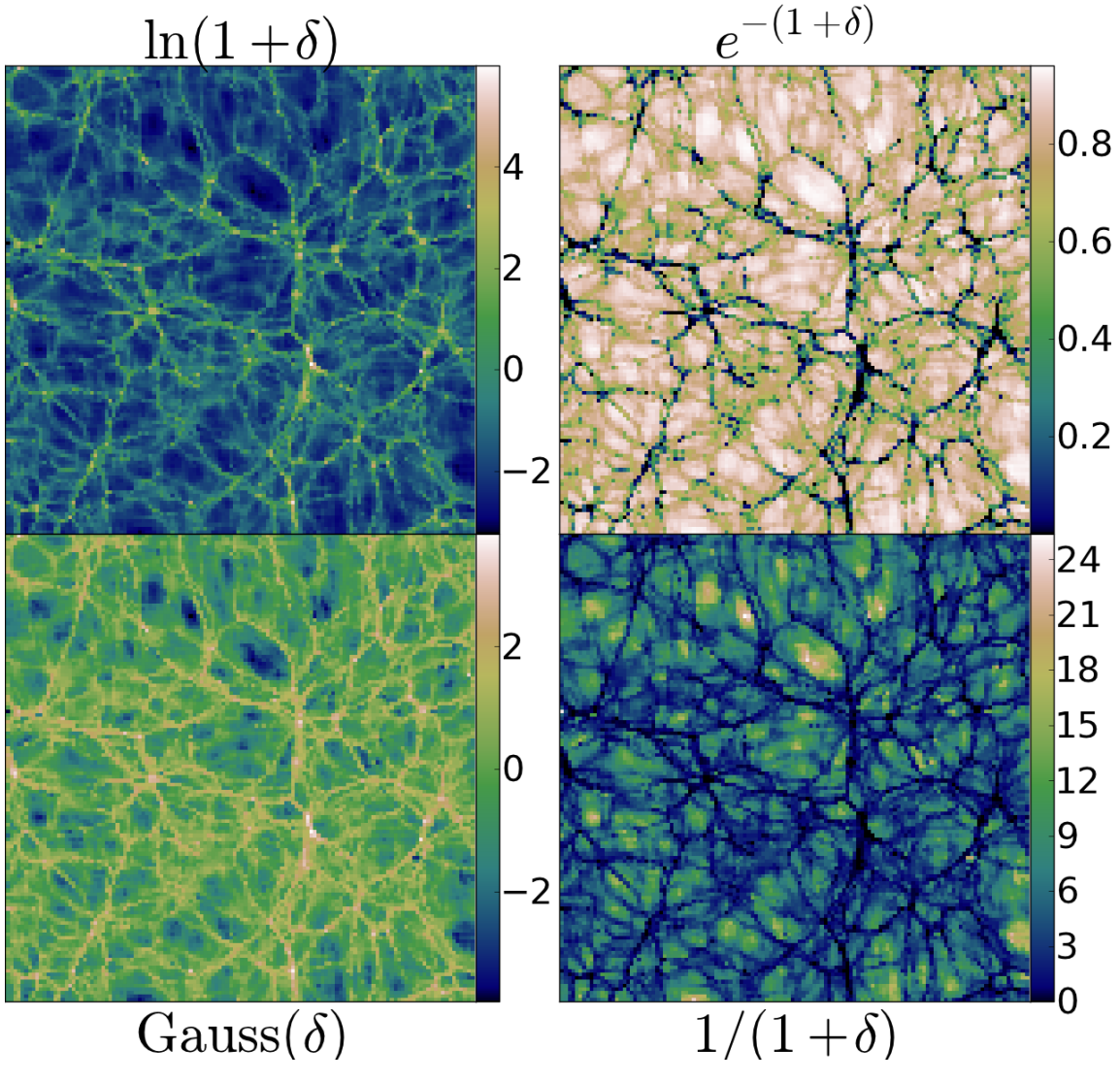

Figure 4.1: Plots of the transformed density-field slices from simulations used. The width of each panel is $128\,h^{-1}$ Mpc, half the box size, and the pixel size is $1\,h^{-1}$ Mpc. The Lagrangian-tessellation density estimate allows accurate density estimates deep within voids, with suppressed particle-discreteness effects compared to other methods.

a cloud-in-cell (CIC) density assignment on a 256^3 grid. The largest wavenumber plotted is that where power spectra estimated with a nearest-grid-point (NGP) method departs noticeably from the CIC power spectra.

However, for $\ln(1+\delta)$, Gauss(δ), and especially $\frac{1}{1+\delta}$, care is required in density estimation because of sensitivity to low-density regions. The results also depend somewhat on the cell size to which the transforms are applied; here, we use $1\,h^{-1}\,$Mpc cells. We use what we call the Lagrangian Tessellation Field Estimator (LTFE),[107,145] referencing the Delaunay and Voronoi Tessellation Field Estimators (DTFE, VTFE),[108] which use Eulerian tessellations. Fig. 4.1 shows transformed-density slices measured this way, from our simulations. The Lagrangian tessellation has a more physical meaning than the Eulerian tessellations, although it has the drawback that it requires knowledge of the ICs, and thus cannot directly be applied to observations. In the LTFE, particles are treated not as mass blobs, but as vertices on a dark-matter sheet (see also[97,146]). For most purposes, and especially for low densities, this approximation is better than treating them as mass blobs, since a physical dark-matter particle is dozens of orders of magnitude lighter than a simulation 'particle.' Lagrangian space is tessellated into tetrahedra according to the initial cubic particle lattice. Matter is uniformly deposited into tetrahedra, which often overlap in high-density regions where streams cross. No pixel is empty, since for each position, there is at least one tetrahedron that gets stretched across it. This method reveals fluctuations in low-density regions that would be difficult to see using a more naive density-estimation method. Much higher mass resolution would be necessary to measure $P_{1/(1+\delta)}$ using CIC.

Ideally, the LTFE would be computed in our cubic cells by computing the intersections

of tetrahedra with each cubic cell. However, for speed, we instead sample the density on a cubic lattice of points. For each tetrahedron enclosing a given lattice point, we add densities inversely proportional to the tetrahedron's volume. To reduce the noise from this point-sampling, each grid point comes from an eightfold density super-sampling; the density is measured on a 512^3 grid, which we then average down to a 256^3 grid. Even so, the process requires much more computation than e.g. CIC, so we implemented a fast GPU code employing CUDA technology.

4.2 Results

Fig. 4.2 shows various power spectra at $z = 0$ from one ensemble of spiked simulations. Also shown is the initial density field used to generate the ICs, $P_{\rm init}$. At large scales, initial spikes are preserved in each power spectrum. But on small scales, each power spectrum behaves differently. At $k \sim 1\,h\,\mathrm{Mpc}^{-1}$, the initial scales dominating P_δ correspond to the green color (larger scales), while the initial scales dominating $P_{1/(1+\delta)}$ correspond to violet (smaller scales).

Figs. 4.3 and 4.4 show the linear-response matrix

$$G_{ij} \equiv \frac{\partial \ln P(k_i)}{\partial \ln P^{\rm init}(k_j)}. \tag{4.1}$$

We use three ensembles of simulations to lower the realization-to-realization variance in the measured G_{ij}, which is typically at the $\sim 10\%$ level, depending on scale. Fixing all mode

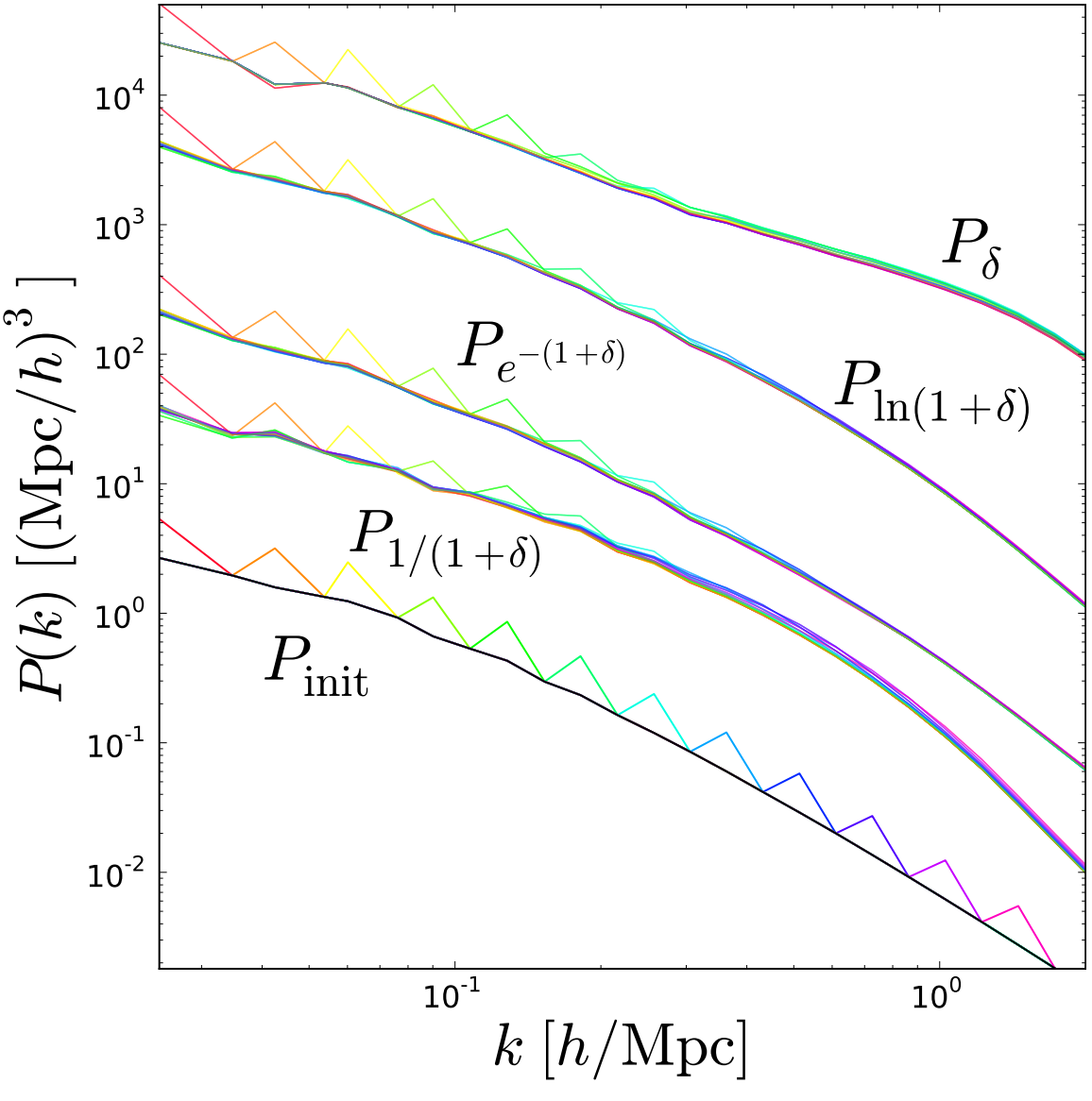

Figure 4.2: From top to bottom, $z = 0$ power spectra from initial-spike simulations of various transformed-density power spectra, along with the spiked initial conditions at $z = 127$. Power spectra are rainbow-colored according to the wavelength of their initial spike, from short (violet) to long (red). $P_{1/(1+\delta)}$ is divided by 10^3 for clarity. $P_{\rm Gauss(\delta)}$, investigated below, is omitted because it is almost indistinguishable from $P_{\ln(1+\delta)}$.

amplitudes to their ensemble-mean values reduces this variance, but differences in random phases still produce fluctuations. The reported value of each G_{ij} matrix element is the median among the three ensembles.

In δ, the migration of power is qualitatively as in Ref.[147] and Ref.:[148] power moves from large to small scales. However, a fixed initial scale does not migrate to a fixed non-linear scale. Translinear ($k \sim 0.3\,h\,\mathrm{Mpc}^{-1}$) initial scales end up smeared over a wide range of smaller scales. The result is that small ($k \sim 1\,h\,\mathrm{Mpc}^{-1}$) final scales are dominated by translinear initial modes, and are insensitive to initial power at their own comoving wavenumber. Indeed, initial power inserted at $k \sim 1\,h\,\mathrm{Mpc}^{-1}$ hardly affects P_δ over the scales measured, although it might show up at smaller scales in higher-resolution simulations. Black curves show an approximation of the $z = 0$ density propagator, a Gaussian with $\sigma_k = 0.2\,h\,\mathrm{Mpc}^{-1}$(CS06). The propagator is a cross-correlation of modes in the initial and final conditions, sensitive to both amplitudes and phases, and, as expected, it follows G^δ_{ij}'s diagonal quite well.

In $\ln(1+\delta)$ and Gauss(δ), the behavior is different. Power spreads out, moving not just from large to small scales, but vice-versa, rather symmetrically. This makes sense: P_δ is mostly sensitive to high-density regions, where fluctuations have contracted. $P_{\ln(1+\delta)}$ and $P_{\mathrm{Gauss}(\delta)}$, on the other hand, are sensitive to dense regions, but also to low-density regions, where fluctuations have expanded. In fact, initial peaks migrate to a bit larger scales than they were initially (visible as slight upturns in Fig. 4.4). There are a couple of possible reasons for this upturn: underdense regions, even if equal by mass, dominate by volume. Also, in overdense regions, many fluctuations completely collapse, perhaps leaving

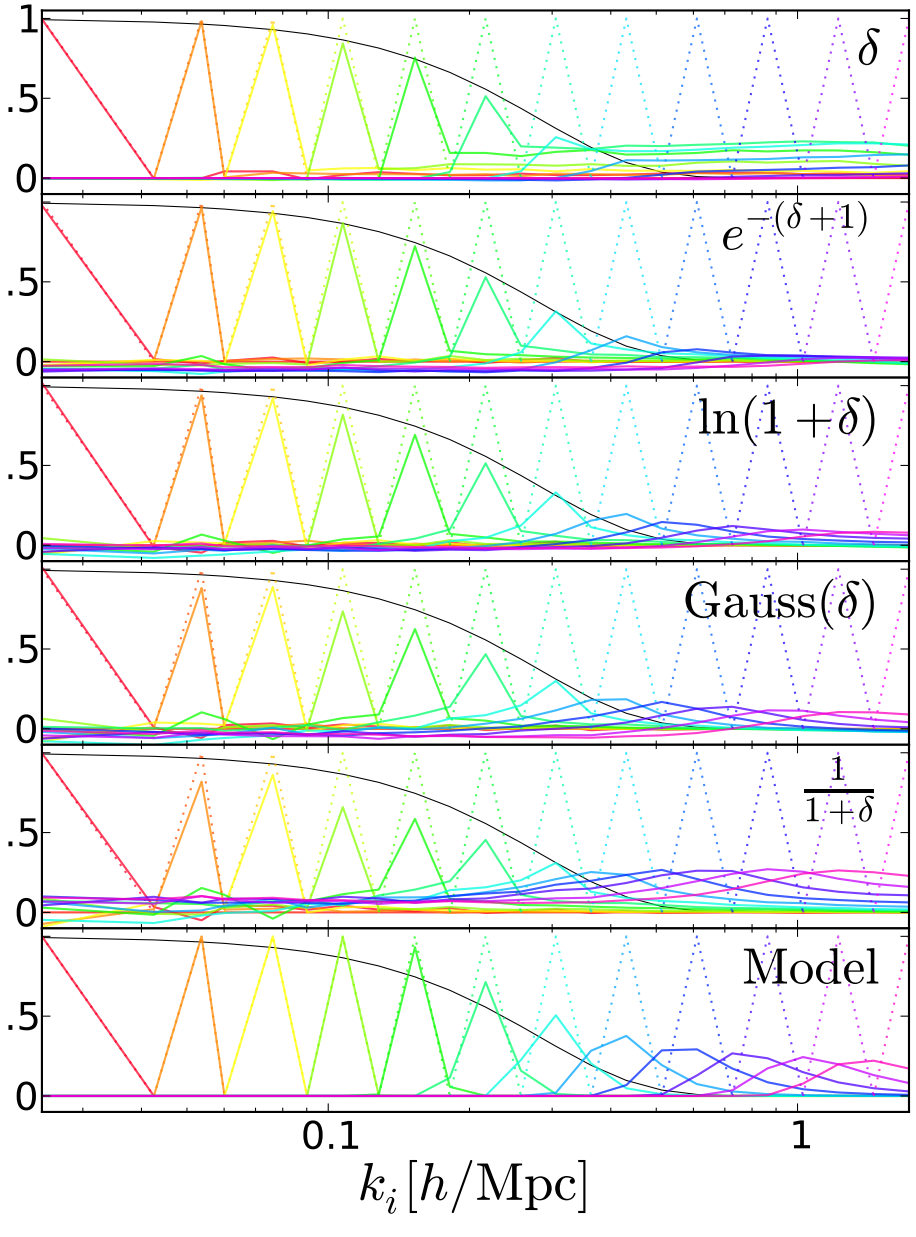

Figure 4.3: Plots of the matrix G_{ij}, defined in Eq. (4.1), showing the response of final-conditions power spectra to initial spikes. The spikes in the initial conditions are shown with dotted curves, rainbow-colored from red to violet going from low to high frequency. Corresponding power spectra of final-conditions simulations appear as solid curves. From top to bottom, the final-conditions resting place of a moderate-scale (e.g. green) spike moves from small to large scales, as each transformation increasingly emphasizes underdense regions. For clarity, power spectra from only odd-numbered spikes are shown. The 'model' is a toy model of power spreading based on a local spherical collapse or expansion of volume elements, given in Eq. (4.4). The black curves show the density propagator (CS06).

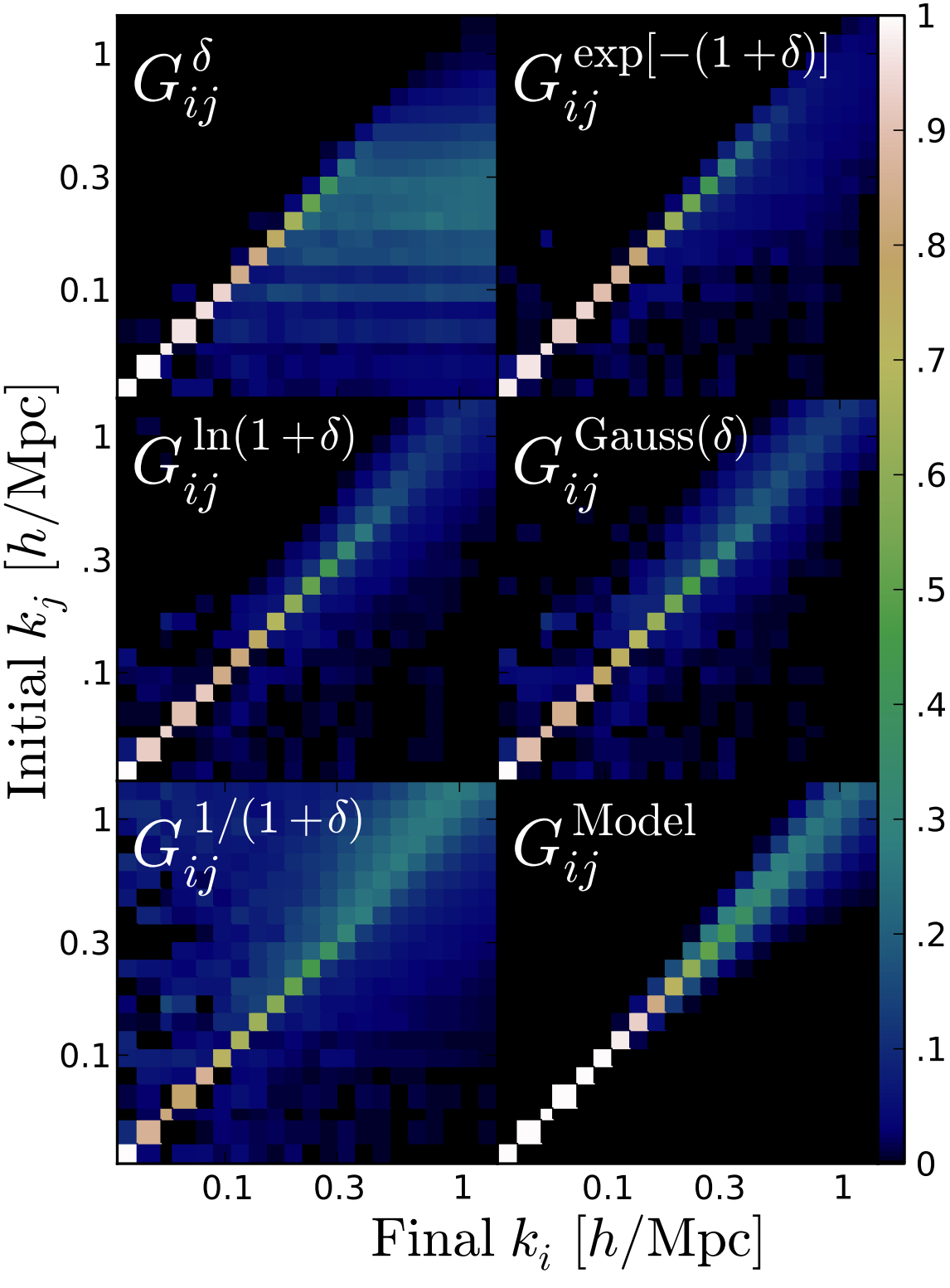

Figure 4.4: G_{ij}'s as in Fig. 4.3, shown in matrix form.

no clear signal in the power spectrum. In underdense final regions, however, almost all fluctuations remain, stretched-out compared to the ICs. Indeed, in the $\frac{1}{1+\delta}$ field, in which underdense regions receive almost all weight, the high value of $G_{ij}^{1/(1+\delta)}$ above the diagonal indicates this comoving expansion of fluctuations. This explicitly confirms the magnifying effect (AS13[117]) that voids have on initially small scales.

Going from δ down to $\frac{1}{1+\delta}$ in Fig. 4.3, the sensitivity to small-scale initial power increases substantially. In the transformed fields, although each individual final mode still retains little memory of its initial phase and amplitude, as quantified by the propagator,[149] small-scale spikes do leave behind obvious bumps at approximately their initial scale. There is a tradeoff on large scales, though: diagonal entries decease slightly, and off-diagonal terms fluctuate increasingly from zero. Large-scale modes, exceeding the scales of typical displacements, are fixed in P_δ by mass conservation, but after a transformation, this is no longer assured, so some fluctuations in large-scale mode amplitudes occur. However, there seems to be no systematic trend in these fluctuations; averaging over many realizations of phases, it seems that these off-diagonal elements would average to zero.

4.3 Toy model for power spreading

A toy model of $G_{ij}^{\mathrm{Gauss}(\delta)}$ based on Lagrangian dynamics captures its behavior qualitatively. The model, which does not explicitly consider the Gaussianization process, is that initial fluctuations expand or contract according to the local density, and that the final $P_{\mathrm{Gauss}(\delta)}$ manages to pick up every fluctuation at the scales where it has expanded or con-

tracted in final conditions. In reality, some initial fluctuations collapse, or otherwise escape detection by $P_{\mathrm{Gauss}(\delta)}$, but the approximation here is that they do not.

Consider fluctuations imprinted on initial pixels of equal Lagrangian size that expand or contract according to the local density. In the ZA, $\nabla_q \cdot \mathbf{\Psi} = -\delta_{\mathrm{lin}}$, where $\mathbf{\Psi} = \mathbf{x} - \mathbf{q}$ is the Lagrangian displacement field, and δ_{lin} is the linearly extrapolated initial density. Approximating each pixel's expansion or contraction as isotropic, with Lagrangian displacement-divergence 'stretching' parameter $\psi \equiv \nabla_q \cdot \mathbf{\Psi}$, each of its 3 dimensions will scale by a factor $1 + \psi/3$. On logarithmic plots such as Figs. 4.3 and 4.4, a fluctuation occupying pixels that stretch in such a way gets shifted by $s \equiv -\ln(1 + \psi/3)$, the minus sign because of the reciprocal from working in Fourier space.

A Gaussian PDF of ψ, with variance σ, can be transformed into a mass-weighted (Lagrangian volume) PDF of s in the ZA:

$$\mathcal{P}(s) = \mathcal{P}(\psi) \left| \frac{d\psi}{ds} \right| = \frac{3 \exp\left[-(3e^{-s} - 1)^2/(2\sigma^2) - s\right]}{\sqrt{2\pi\sigma^2}}, \tag{4.2}$$

where $\mathcal{P}$ denotes a probability distribution.

Alternatively, a spherical-collapse approximation can be used for the behavior of ψ, where ψ_{z} is its ZA value,[103,150]

$$\psi_{\mathrm{sc}} = 3 \left[\left(1 + \frac{2}{3} \psi_{\mathrm{z}} \right)^{1/2} - 1 \right]. \tag{4.3}$$

This uses an approximation for the evolution of a volume element found by Ref.,[101] valid

in the limit of low matter density. It gives the following distribution of s:

$$\mathcal{P}(s) = \frac{(3/2)\exp\left[-(3/2)(e^{-2s}-1)^2/(2\sigma^2) - 2s\right]}{\sqrt{2\pi\sigma^2}} \tag{4.4}$$

Both Eqs. (4.2) and (4.4) can be approximated by a Gaussian of dispersion $\sigma/3$ for small σ.

The final panels of Figs. 4.3 and 4.4 show G_{ij} in this model. The assumption is that Eq. (4.4) gives the shapes of the curves into which spikes broaden in Fig. 4.3. Each j row is normalized so that all initial fluctuations contribute to some final wavenumber, i.e. so that for all j, $\Sigma_i G_{ij}^{\rm Model} = 1$. Physically, this would describe G_{ij} for a density variable that manages to capture all Lagrangian volumes, both expanding and contracting. It is not surprising that $G_{ij}^{\rm Model}$ has higher amplitude on small scales than any measured G_{ij}, since some fluctuations doubtless evade notice by any final power spectrum, i.e. $\Sigma_i G_{ij} < 1$ generally. We estimate σ^2 in Eq. (4.4) as the variance in top-hat spheres of radius $2\pi/k_j$ in linear theory. This model is quite naive; for instance, it assumes an equal-Lagrangian-volume (i.e. mass) weighting, not an Eulerian equal-volume-weighting. Still, the model captures the qualitative behavior of $G_{ij}^{{\rm Gauss}(\delta)}$, although it may underestimate the variance at low k_j, and overestimates $G_{ij}^{{\rm Gauss}(\delta)}$ along the diagonal.

4.4 Covariance Matrices and Information

G_{ij} can be used to estimate power-spectrum covariances, as well. Usually, covariance matrices are measured directly from ensembles of final density fields, and it is interesting to compare this approach with the result of directly tracking initial-conditions degrees of freedom. A linear model of a fluctuation away from the mean in the final-conditions $\ln P_i$ (investigated instead of P_i for algebraic simplicity) is, summing over repeated indices,

$$\Delta \ln P_i = \frac{\partial \ln P_i}{\partial \ln P_j^{\rm init}} \Delta \ln P_j^{\rm init} = G_{ij} \Delta \ln P_j^{\rm init}, \tag{4.5}$$

giving

$$C_{ij} = \langle \Delta \ln P_i \Delta \ln P_j \rangle = \left\langle G_{ik} \Delta \ln P_k^{\rm init} G_{jl} \Delta \ln P_l^{\rm init} \right\rangle . \tag{4.6}$$

Since the Gaussian $C_{ij}^{\rm init} \equiv \left\langle \Delta \ln P_i^{\rm init} \Delta \ln P_j^{\rm init} \right\rangle = 2\delta_{ij}^{\rm K}/N_i$, where N_i is the number of modes in bin i,

$$\mathsf{C} = \mathsf{G}\mathsf{C}^{\mathsf{init}}\mathsf{G}^{\top}. \tag{4.7}$$

Now, suppose we want to estimate all $\ln P_i^{\rm init}$ from the final power spectrum. The Fisher matrix F_{ij} to use to predict constraints on this initial power in bins i and j would be

$$F_{ij} = \frac{\partial \ln P_k}{\partial \ln P_i^{\rm init}} (\mathsf{C}^{-1})_{kl} \frac{\partial \ln P_l}{\partial \ln P_j^{\rm init}} = G_{ki} (\mathsf{C}^{-1})_{kl} G_{lj}, \tag{4.8}$$

or

$$\mathsf{F} = \mathsf{G}^{\top} (\mathsf{G}\mathsf{C}^{\mathsf{init}}\mathsf{G}^{\top})^{-1} \mathsf{G} = (\mathsf{C}^{\mathsf{init}})^{-1}. \tag{4.9}$$

So the covariance matrix of parameters that consist of the initial power spectrum in bins is just the initial, Gaussian power-spectrum covariance matrix. This suggests no Fisher-information loss!

However, this calculation assumes that the final power spectrum is an entirely deterministic, invertible, linear transformation of the initial power spectrum, with no sources of noise. This is not the case; it neglects at least a couple of things: the non-linear coupling of pairs of power-spectrum spikes, and realization-to-realization fluctuations in the G_{ij} matrix, which can depend on both the power spectrum itself, and also on mode correlations (present even in a Gaussian field) that affect the halo mass function, which substantially affects translinear-scale power, at least in the halo model.[128]

One neglected factor that can be investigated in a straightforward extension of the present framework is the non-linear coupling of spike pairs. But this would involve a simulation for each pair of wavenumber bins, i.e. with 23 bins, $23 \times 22 = 506$ additional simulations. We plan to run this brute-force ensemble in future work. For now, we compare the covariance matrix from Eq. (4.5) to one estimated otherwise.

Fig. 4.5 shows the non-Gaussian part of the δ power-spectrum covariance $T^{\delta}_{ij} \equiv C^{\delta}_{ij}(N_i N_j)^{1/2} - \delta^{\rm K}_{ij}$, both in the linear-response model from G^{δ}_{ij} in Eq. (4.6), and from the fluctuating-multiplicative-bias model of Ref.,[2] a rather accurate approximation to the covariance as measured from the Coyote Universe simulations.[151] In this model, the non-Gaussian covariance is given by $T^{\delta}_{ij} = \alpha(N_i N_j)^{1/2}$, where α is the fractional realization-to-realization variance of the nonlinear density-field variance in nonlinear-scale cells. We use $\alpha = 0.0035$

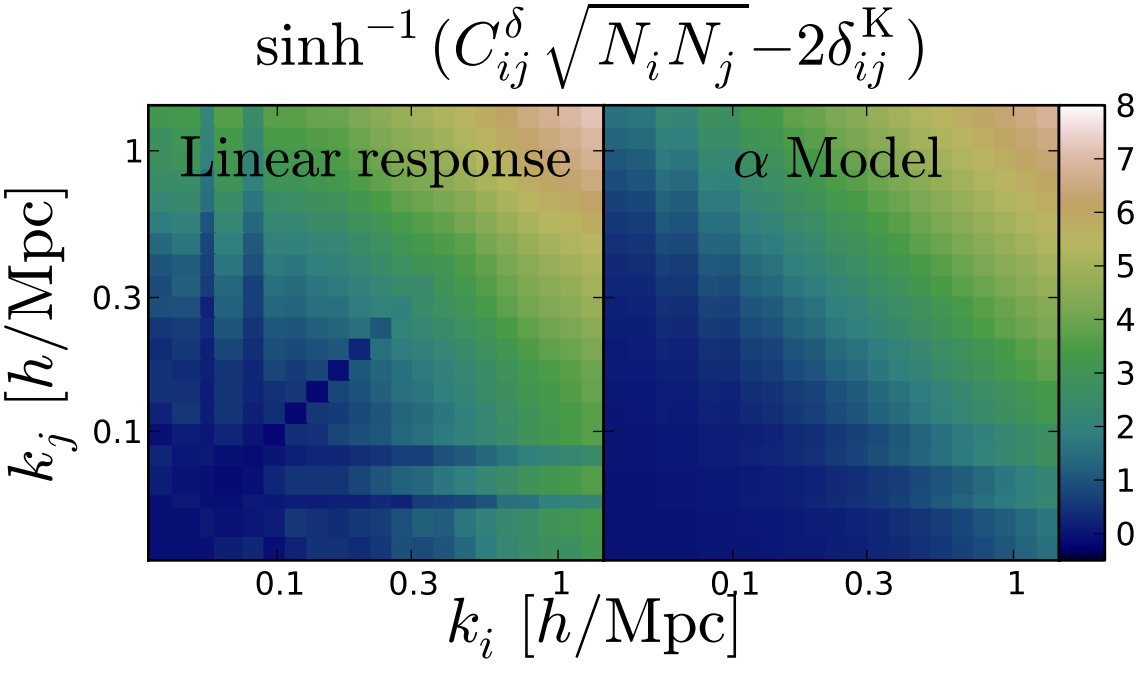

Figure 4.5: The non-Gaussian part of the δ power-spectrum covariance T_{ij}, as measured from G^{δ}_{ij} using the linear-response model of Eq. (4.6), and using an 'α model' approximation to T_{ij} found in.[2] Because of the steep increase, we use a $\sinh^{-1}$ transform for plotting, which becomes logarithmic for large values of its argument. While the α model should not be taken too seriously as it is only an approximation, the qualitative agreement (except, perhaps, far from the diagonal) between the two plots suggests that the linear-response model captures most of the relevant effects.

for this plot, which is α at $z = 0$ as found by Ref.,[2] scaled to the $(256h^{-1}\,\mathrm{Mpc})^3$ volume of the present simulations.

We emphasize that the α model is approximate, but qualitatively, it agrees with the linear-response model rather well, suggesting that additional terms in the covariance may indeed be subdominant. The main discrepancies are in highly off-diagonal terms.

4.5 Conclusions

We used an N-body experiment to track where initial power-spectrum features get deposited in final-conditions density power spectra. For the usual overdensity field δ, our results qualitatively agree with the common wisdom that initial power migrates from large to small scales. However, this seems to be largely because the δ field is dominated by overdense spikes. When the density is transformed to have a more-Gaussian PDF, increasing the statistical weight of low-density regions (where patches imprinted with initial fluctuations expand rather than contract in comoving coordinates), initial spikes spread rather symmetrically, both upward and downward in scale. In fact, in $P_{1/(1+\delta)}$, almost exclusively sensitive to underdense regions, initially small scales are magnified (AS13). In these power spectra, initial small-scale spikes leave much more evidence at $z = 0$ than in P_δ.

The spread of power in the Gaussianized variables such as $\ln(1 + \delta)$ is qualitatively captured by a toy model we give, in which patches imprinted with initial fluctuations expand or contract according to a spherical-collapse model. In the future, it would be interesting to refine this model for greater accuracy, and investigate whether it might be modified

successfully to other power spectra.

We also begin to apply our results to the theoretical question of how degrees of freedom present in the initial density field, essentially a sum of many spikes such as the ones we use, disappear from the final-conditions, coarse-grained density field. However, this will require further measurements, because in our framework, fluctuations in the final power spectrum are a linear, invertible transformation of the initial power spectrum, given by a matrix G_{ij}. In reality, though, information is lost because of a few neglected effects, which we will analyze in future work. We do, however, find that G_{ij} gives a rather accurate description of the P_δ covariance matrix, suggesting that one of these effects (the nonlinear coupling of spike pairs) is not dominant. Thus, 'ringing' the initial power spectrum as we do offers an interesting technique to estimate power-spectrum covariances and Fisher information, by tracking the true, initial degrees of freedom in the Universe.

Chapter 5

Dark Matter Halo Spin Alignment Under CDM

This chapter is based on Aragon-Calvo & Yang (2014).[152]

In the gravitational instability scenario galaxies acquire their angular momentum via tidal torque produced by a misalignment between the inertia tensor of the proto-halo and the embedding gravitational shear tensor generated by the surrounding matter distribution. This is the so called Tidal Torque Theory (TTT).[153,154] Although the TTT predicts a correlation between the spin orientation of galaxies and their local Large Scale Structure (LSS) observations of spin alignments have been so far inconclusive and often contradictory (Ref.[155–162] among others). The reason is still not clear and may include the intrinsic difficulties in deducing the spin vector of galaxies, small galaxy samples, challenges in characterizing LSS from galaxy redshift catalogues, etc. This is in adition to complex

galaxy formation processes such as anisotropic infall and mergers which may lower or even remove any priordial spin alignments.

On the theoretical/numerical front the picture is more clear. To date there is a consensus that the spin vector of haloes in walls lies in the plane of their host walls while haloes in filaments have their spin either parallel or perpendicular to their host filament depending on their mass.[163–168] An important aspect of the spin alignment of haloes in filaments is the existence of a transition mass M_{tr}, at which haloes change their orientation from parallel to perpendicular. This effect was first reported by Ref.[163] and later confirmed by other authors.[164–168] The spin alignment of low-mass haloes also has a redshift dependence starting perpendicular to their host filament and later becoming parallel around $z \sim 1$.[167,169] The change in the primordial orientation indicates an additional mechanism for angular momentum acquisition that has a stronger effect in low-mass haloes.

The existence of a hierarchy of structures in the Cosmic Web in which large structures contain and are defined by smaller ones (Ref.[117,170,171] and references therein) raises the question of whether the halo alignment reflects this hierarchy. The limited number of studies that link the alignment signal and the hierarchy of cosmic structures indicate that the transition mass of alignment in filaments depends on the scale of the filament[166] and the distance from the halo to the closest cluster.[165] However, to date no direct link has been stablished between halo alignment and the hierarchy of cosmic structures. In the observational front, a strong perpendicular alignment of low-luminosity small galaxies in tenuous intra-wall filaments has been observed by Ref.[172] The two regimes of alignment/anti-alignment of low-mass and high-mass haloes in filaments have been observed by Ref.[162] if one assumes

a correlation between Hubble type and galaxy/halo mass. They found that bright spirals have a weak tendency to be parallel to their host filaments while Elliptical/S0 show a strong perpendicular alignment.

The TTT theory does not provide a mechanism (at least a simple one) to produce the observed mass-dependent two-regime alignment of haloes or to change the alignment of haloes from parallel to perpendicular. An alternative mechanism for angular momentum acquisition was presented by Ref.[173] and more recently discussed by Ref.[166,174–176] and Ref.,[167] in which angular momentum is generated by the anisotropic infall of matter. This is in turn defined by the surrounding LSS morphology and dynamics thus offering a natural process for spin-LSS correlations. This secondary-torque mechanism may help explain the change in alignment and possibly even the primordial torque generation.

5.1 N-body Simulation and DM haloes: the MIP correlated ensemble

The N-body simulations and related halo catalogues used in this work are based on the MIP correlated ensemble simulation.[144] The MIP simulation consists of 220 realizations (at the time of submision). Each realization in the ensemble contains 256^3 dark matter particles inside a box of 32 Mpc h^{-1} of side with standard cosmological parameters. All realizations in the ensemble share identical LSS modes (defined at $k < k_{\rm cut}$). Where $k_{\rm cut}$ corresponds to a scale of 4 Mpc h^{-1}. At smaller scales ($k > k_{\rm cut}$) realizations can be considered independent see Ref.[144] for details. The MIP simulation was designed to address a variety of problems

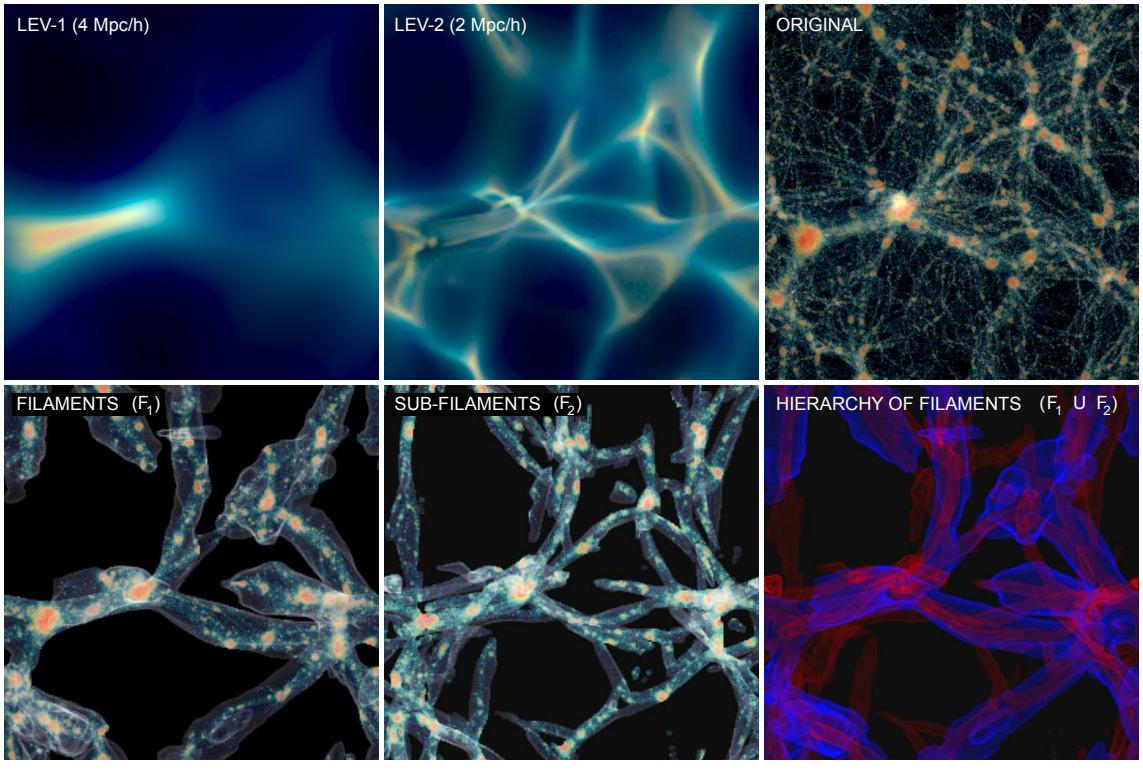

Figure 5.1: The MMF-2 method. Top panels: hierarchical space. We show the volume rendering of the density field at $z = 0$ for linear-regime smoothing at 4, 2 Mpc h^{-1} and no smoothing (left, center and rights panels respectively). The thickness of the slices was chosen to show as many structures as possible while avoiding confusion. Bottom panels: Hierarchical identification of filaments. The left and center panels show the density field inside filaments (F_1) and sub-filaments (F_2) respectively. We highlight the filament mask as a semitransparent surface. The hierarchy of filaments (blue) and sub-filaments (red) is shown in the bottom right panel.

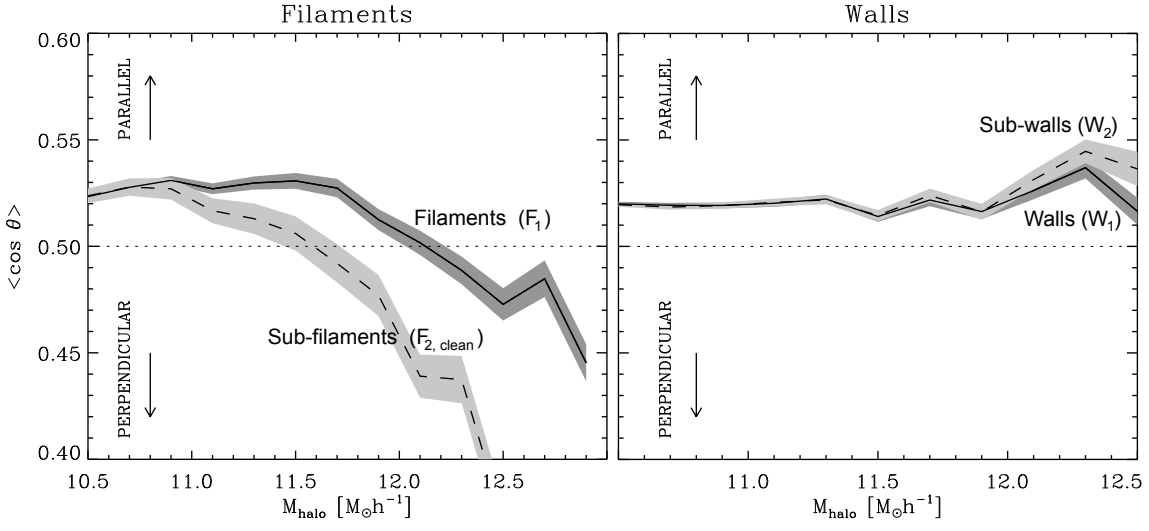

Figure 5.2: Spin alignment of haloes as function of mass in filaments (left panel) and walls (right panel). The alignment is measured with the mean $\cos\theta$ where 0.5 corresponds to no preferred alignment. The sample was divided in haloes in filaments/walls (solid lines) and sub-filaments/sub-walls (dashed lines). Error bars per bin were computed from 1000 random realizations with the same number of points as their corresponding bin.

including increased signal to noise ratio for halo statistics in low-density environments,[177] local studies of scalar and vector fields[178] and local ensemble statistics such as local halo mass function.[144] In the present study we take advantage of the common LSS between realizations of the MIP allowing us to perform LSS characterization at high spatial resolution while including a large number of haloes.

5.1.1 Halos

Haloes were identified in each realization of the MIP ensemble using the friends-of-friends algorithm with a linking parameter $b = 0.2$. For this study we use haloes defined by 50 or more particles corresponding to a minimum halo mass of $\sim 8\times10^9$ $\mathrm{M}_\odot h^{-1}$. We use the MIP ensemble in "stack mode", where halo catalogues from all realizations are aggregated

into one "master halo catalogue". The full MIP stacked ensemble contains $\sim 5 \times 10^6$ haloes inside a volume of $(32\ \mathrm{Mpc\ h^{-1}})^3$ giving unprecedented halo number density. For each FoF group, we compute its mass, center of mass, mean velocity etc. The angular momentum was computed as $\mathbf{J} = \sum_i^N m_i\ \ \mathbf{r_i} \times (\mathbf{v_i} - \bar{\mathbf{v}})$. Where the sum is over all the particles in the halo, m_i is the particle mass, $\mathbf{r_i}$ is the distance to each particle from the center of the halo, $\mathbf{v_i}$ is the peculiar velocity of the particle and $\bar{\mathbf{v}}$ the mean velocity of the halo with respect to its center of mass.

5.2 MMF-2: Hierarchical Identification of Filaments and Walls

In order to study the effect of environment in the alignment of haloes in filaments and walls we need to characterize the density field according to its local geometry, compute its local direction and, for the study presented here, its *hierarchical relations*. The identification of filaments and walls was performed using a significantly improved implementation of the Multiscale Morphology Filter (MMF, Ref.[179]).

The identification of structures in the original MMF method and the one presented here (MMF-2) is done based on the second-order local variations of the density field encoded in the Hessian matrix ($\partial^2\rho/\partial x_i \partial x_j$). LSS morphologies are associated to ratios between the eigenvalues of the Hessian matrix ($\lambda_1 < \lambda_2 < \lambda_3$). Different ratios measure local spherical symmetry, filamentariness, or planarity. In a filament the density field is relatively constant along its lenght and so $\boldsymbol{\lambda_1}$ points in the direction of the filament. In a wall the largest variation is in the direction perpendicular to the plane of the wall and is traced by $\boldsymbol{\lambda_3}$ (see

Ref.[163] for details on the MMF method and morphology filters).

The MMF-2 performs the identification of structures on hierarchical space in contrast to the scale-space approach used in the orginal MMF and its derived implementations.[165,180] A Hierarchical space is a generalization of scale-space which exposes the *hierarchical character* of the density field in contrast to the scale-space approach which emphasize the *scale* of the structures and is insensitive to their nesting relations. The hierarchical space used in the MMF-2 is able to describe both scale and nesting relations.

The output of the MMF-2 is a set of "masks" sampled on a regular grid indicating which voxels belong to a given morphology at a given hierarchical level. These masks roughly correspond to the "feature maps" $\mathcal{F}$ in the original MMF algorithm. A complete description of the MMF-2 method will be presented in Aragon-Calvo (2013) (in preparation).

5.2.1 Hierarchical space

We generate the hierarchical space neeed for the MMF-2 analysis as described in Ref.[118] and Ref.,[144] by Gaussian-smoothing the *initial conditions* instead of the *final* density field as in the scale-space approach. We refer to this as *linear-regime smoothing* since the filtering is applied when all Fourier modes are independent, allowing us to target specific scales in the density field before non-linearity and Fourier mode-mixing occurs. The smooth initial conditions evolve by gravity producing a density field with all the anisotropic features of the Cosmic Web but lacking of small-scale structures including haloes (below the mass corresponding to the smoothing scale). This reduces the dynamic range in the final density field

and greatly limits the contamination in the identification of filaments and walls produced by dense haloes .

For the purpose of this paper we generate a two-level hierarchical space defined by linear-regime scales of 4 Mpc h^{-1} delineating the large structures in the LSS and 2 Mpc h^{-1} for the smaller sub-structure (see section 5.2). We denote these hierarchical levels as LEV-1 (filaments/walls) and LEV-2 (sub-filaments/sub-walls). From the template initial conditions used to generate the MIP ensemble (see Ref.[144] for details) we generated low-resolution 128^3 initial conditions grids smoothed at the two scales defined above. We then evolved both smoothed simulations to the present time $z = 0$ using the PM-Tree code GADGET-2.[98] It is important to emphasize that these two simulations were produced only in order to generate density fields for the MMF-2 LSS hierarchical characterization.

5.2.2 Density fields

From the particle distribution we computed densities following the method presented in Ref.[107] In this novel approach particles define Lagrangian volumes that reflect the changes in density as particles evolve deforming the original volume. Lagrangian volumes are defined by 8 adjacent particles forming a "cube" that is divided into 6 tetrahedra. Densities are estimated in the tessellation as the mean density of the adjacent tetrahedra to each vertex and interpolated to a regular grid taking care to account for multistreams. This method makes density estimation computationally straightforward, it produces a continuous volume-filling density field with practically no artifacts in low and medium-density regions (our regions of interest) and is self-adaptive by construction.

5.2.3 The filament/sub-filament sample

From the density fields we then proceed to do the LSS characterization using the MMF-2 as described in section 5.2. We generated filament, sub-filament wall and sub-wall masks (F_1, F_2, W_1 and W_2 respectively, the subscript here indicating the hierarchical level). Figure 5.1 shows (top panels from left to right) the density field at $z = 0$ from the two hierarchical levels LEV-1 and LEV-2 (see section 5.2.1) and the original initial conditions (no linear regime smoothing). The original density field enclosed in the filament and sub-filament mask is shown at the bottom-left and center panels. The two-level hierarchy of filaments (the union of the two sets) is shown in the bottom-right panel.

In order to clearly differentiate between the alignment signal of haloes in filaments and sub-filaments we differentiated between "isolated" sub-filaments and embedded sub-filaments (being inside filaments). We additionally constrained sub-filaments to be contained by walls in LEV-1 in order to remove tenuous intra-void filaments. We define the mask of "clean" sub-filaments as:

$$F_{2,\mathrm{clean}} = F_2 \cap W_1 \cap F_1^{\mathrm{C}} \tag{5.1}$$

where F_2 is the sub-filament mask, W_1 is the wall mask and F_1^{C} is the complement of the filament mask. In what follows F_1 and $F_{2,\mathrm{clean}}$ denote the masks we used to compute spin-alignment of haloes in filaments and sub-filaments respectively.

5.2.4 Spin alignment in filaments and walls

The angle between the spin of the halo and its host filament or wall was computed as $\theta_F = \phi_W$ and $\theta_W = 90° - \phi_W$ respectively, where $\phi_F = \cos^{-1}((|\mathbf{J} \cdot \lambda_\mathbf{1}|)/(|\mathbf{J}| \cdot |\lambda_\mathbf{1}|))$ and $\phi_W = \cos^{-1}((|\mathbf{J} \cdot \lambda_\mathbf{3}|)/(|\mathbf{J}| \cdot |\lambda_\mathbf{3}|))$ are the angles between the spin of the halo the and direction of its host filament ($\lambda_\mathbf{1}$ or wall ($\lambda_\mathbf{3}$).

5.3 Results

Our main findings are shown in Figure 5.2 where we measure the mean cosine of the angle between the spin vector of haloes and the direction of their host filament/sub-filament (left panel) or wall/sub-wall (right panel) as a function of halo mass. A more detailed view of the spin alignment distribution of haloes in filaments/sub-filaments is shown in Figure 5.3. Our results can be summarized as follows:

- Haloes in filaments and sub-filaments present a two-regime spin alignment as function of their mass: low-mass haloes are (on average) aligned parallel to their parent filament while high-mass haloes are oriented perpendicular. This confirms previous findings.

- The mean alignment of low-mass haloes seems to converge to around $< \cos\theta >= 0.53$ down to a halo mass of $\sim 10^{10}$ $\mathrm{M}_\odot$ h^{-1} in both filaments and sub-filaments.

- The alignment of high-mass haloes does not seem to converge like in the case of the low-mass haloes. Instead, the alignment signal increases with increasing mass.

- We found a significant difference in the transition mass M_{tr} dividing parallel and perpendicular alignment for haloes in filaments and sub-filaments being $M_{tr,low} \sim 1.5 \times 10^{12}$ $M_\odot$ h^{-1} and $M_{tr,high} \sim 4 \times 10^{11}$ $M_\odot$ h^{-1} respectively.

- The slope in the alignment-mass curve of high-mass haloes in sub-filaments is more pronounced than in the case of haloes in filaments.

- Haloes in walls/sub-walls have their spin oriented in the plane of the wall for the full range of masses in our halo sample ($\sim 10^{10} - 10^{12.5}$ $M_\odot$ h^{-1}).

- While the spin alignment distribution of haloes in filaments and sub-filaments seem scaled versions of each other (see Figure 5.3) the spin alignment distribution of low-mass haloes is practically the same for filaments and sub-filaments even though their transition masses are different by a factor of 3.

5.4 Conclusions and discussion

Using an advanced LSS classification algorithm (MMF-2) and a novel N-body correlated ensemble simulation we were able to measure the halo spin alignment signal with unprecedented detail in a variety of cosmic environments. Our results confirm previous findings while significantly improving the alignment signal detection. We found that the spin alignment of haloes in filaments has a hierarchical character reflected in both the strength of the alignment and the transition mass M_{tr} at which the spin alignment changes from being parallel (low-mass haloes) to perpendicular (high-mass haloes). This hierarchy-dependent

effect has not been observed before due to the inability to explicitly target the nesting relations of the structures in the Cosmic Web.

We found that haloes in walls have their spine aligned with the plane of the wall independently of the halo's mass.

The explicit link between halo spin alignment and Cosmic Web hierarchy was possible with the use of the MMF-2 method (introduced here) that identifies cosmic structures and their hierarchical relations and benefits from an improved density estimation algorithm.

The use of the MIP simulation allowed us to perform a high-resolution analysis of the density field over a relatively small volume while at the same time providing us with the large number of haloes needed to detect the spin alignment signal.

While the transition mass for haloes in filaments/sub-filaments may be affected by the choice of the scales defining the hierarchical levels the observed trend should still be valid for other hierarchical levels.

The box size is relatively small for a cosmological simulation. This may introduce unwanted effects from the lack of large modes and non-linear effects from self-interaction of the LSS. These effects can have a modulating effect in the results presented here, giving a more qualitative view of the effect we are studying rather than an stricktly quantitative one.

5.4.1 On the origin and fate of the spin alignment

The change of alignment of haloes in filaments, the dependence of the transition mass with redshift[166] and filament hierarchy (this work) supports the idea of a "secondary torque" process that acts cumulatively and affects all haloes but is manifiested first in low-mass haloes. Since tidal torque is most efficient before turnaround the process changing the spin orientation of haloes must be of a different nature. The anisotropic infall scenario offers a tantalizing alternative. Walls and sub-walls accrete mass from their adjacent voids in the direction parallel to the normal of the wall inducing angular momentum on the plane of the wall so haloes retain their primordial spin alignment. On the other hand filaments and sub-filaments accrete mass mainly from their adjacent walls in the direction perpendicular to the filament thus inducing a parallel spin alignment. This anisotropic accretion has a stronger effect in low-mass haloes than in more massive ones. The stronger perpendicular alignment of haloes in sub-filaments in comparison to haloes in filaments reflects the lower density and lower small-scale peculiar velocities characteristic of walls and their embedded sub-filaments. This is a natural consequence of the hierarchical nature of the Cosmic Web.

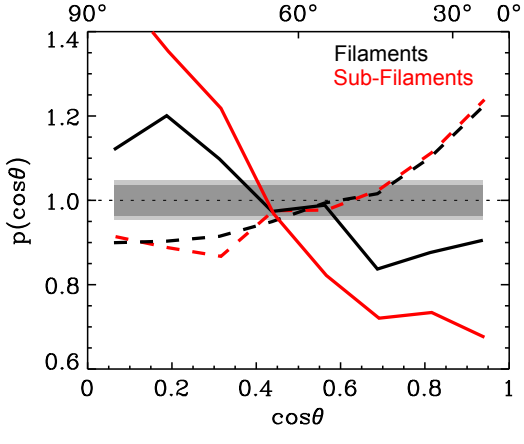

Figure 5.3: Distribution of angles between the halo's spin and its parent filament (black line) or sub-filament (red line). The samples were divided as high-mass haloes (solid line) and low-mass haloes (dashed line). The transition mass for filaments and sub-filaments corresponds to $\sim 1.5 \times 10^{12}$ and $\sim 4 \times 10^{11}$ $M_\odot$ h^{-1} respectively.

Chapter 6

Simulation Data Analyzing Methods

Cosmological data sets usually contain more than 10^9 particles. However, to compare the simulation results with observations, it is essential to run statistical estimations on these datasets. Simulation statistics also provide testers for phenomenological theories. It is usually highly non-trivial to run useful analytics on big data sets. For example, a single calculation of the dark matter annihilation/decay signal on a 10^8-particle simulation might require 100 of CPU hours, which might well be a supercomputer-task. In this chapter, we present three techniques that made our analysis possible without much computational resource.

The first one is the dark matter annihilation/decay signal estimation. The detail technique is presented in Section 6.1. In a simulation, the emission of each tracer particle can

be approximated by a Gaussian profile. The difficulty of obtaining a full sky observation is to sum over all the particles and their Gaussian profiles projected on to the celestial sphere. A direct calculation is almost intractable when the number of particles is too many, e.g., roughly 10^9 of them. We introduce a novel projection, the stereographic projection, which projects a circular shape on the sphere to a circular shape on the 2D plane. This projection further enables us to use GPU to do the rendering from the hardware level. This code allows an 8-hr task (on a single-CPU workstation) to be done within a minute.

The second one is the Lagrangian Tessellation Field Estimator. This code is aiming to compute the density field of a large scale simulation. It is known that density field is essential to analyze the dark matter kinematics. Previously known density estimation methods are either too slow or not sufficiently accurate. Our methods use the so-called Lagrangian Tessellation Field Estimation method that treats each tetrahedron of 4 neighbor particles as a uniform density region. This method gives highly accurate and also physical meaningful density estimation. However, it is highly computation-resource consuming. Using GPU, we construct a novel method that first cut the tetrahedron into slices, where each slice consists one or two triangles. These triangles can be rendered using GPU hardware very efficiently. More details of this algorithm are presented in Chapter 6.2.

The third one is a streaming algorithm that finds the halos in a cosmological data set. The first step of studying the statistics of dark matter halos is to find them. Conventional methods of halo finding usually utilize the so-called friends-of-friends method – a particle and all its neighbors within a distance are included in a halo. By setting a threshold, this method repeatedly identifies all particles that belong to a single halo. As one can

easily see, this method is highly memory and time-consuming. What we propose is a one-pass algorithm, that simply reads through the entire dataset and outputs the positions of the possible halos. Our algorithm exploits the recent progress in streaming algorithms from the theoretical computer science community. Our algorithm made it possible for a supercomputer task to be run on a personal laptop. More details of this algorithm are presented in Chapter 6.3.

6.1 Dark Matter Annihilation Calculation

In a previous study Ref.,[3] the whole sky map of dark matter annihilation Gamma-Ray signal is estimated from the VIALACTEA II (VL-II) simulation to simulate Fermi-LAT's observation. The snapshot used is a 40 Mpc box with about 400 million particles. With an SPH kernel smoothing and certain approximation, each particle's cross-section-independent flux distribution is given by

$$F_i(\theta,\phi)\mathrm{d}\Omega = \begin{cases} N\dfrac{m_i\rho_i}{4\pi d_i^2}\exp\left(-\dfrac{3}{2}\dfrac{\delta\theta^2}{r_\theta^2}\right)\mathrm{d}\Omega & \delta\theta \le r_\theta \\ 0 & \delta\theta > r_\theta \end{cases}, \tag{6.1}$$

where $\delta\theta$ is the angular distance from (θ,ϕ) to the particles angular position; r_θ is the angular radius of the particle; ρ_i, m_i, d_i are the density, mass and distance of the i-th particle. With the help of HEALPIX,[49] each particle's contribution computed as a disk of pixels by Equation (6.1). Calculating the pixel disk is a highly inefficient process. Even though it is embarrassingly parallel to run each particle in a single GPU thread, the direct

CUDA version of the CPU code is even slower. The reason is that the pixel distribution of each particle is highly varied (each particle has average 68 pixels on its HEALPIX disk, but with a standard deviation of 115; moreover, 20k particles have more than 10k pixels). Sorting the particles based on the angular radius could largely reduce the threads skew. However, a highly optimized CUDA code needs more than one hr to run (GPU: GTX 480). In contrast, the c++ code requires about 7 hrs to finish (on a $\sim$3 GHz CPU).

In this chapter, we propose a new method which highly speeds up the calculation. Because of the high inefficiency of querying HEALPix pixel disks, we use the stereographic projection (STR) to project the profile on to a regular easy-to-reach pixel table. The computation of the projection on the regular pixel table can be computed in a GPU with hardware acceleration, which gives us the speed up.

In Section 6.1.1, we introduces the method we used; Section 6.1.2 shows the comparison between the new result and old result; Section 6.1.3 introduces some potential usage of the new method.

6.1.1 GPU Based STR Projection (GSP) Method

STR projection projects a circle on the sphere to a circle on the projection plane. The center is projected to an off-center point. Therefore, the Gaussian profile of a particle's flux is projected to a stretched Gaussian (see Fig-6.1).

With STR, we use OpenGL's Point Sprite feature to render the stretched Gaussians to squared billboards and add them together on the projection plane. STR projection has

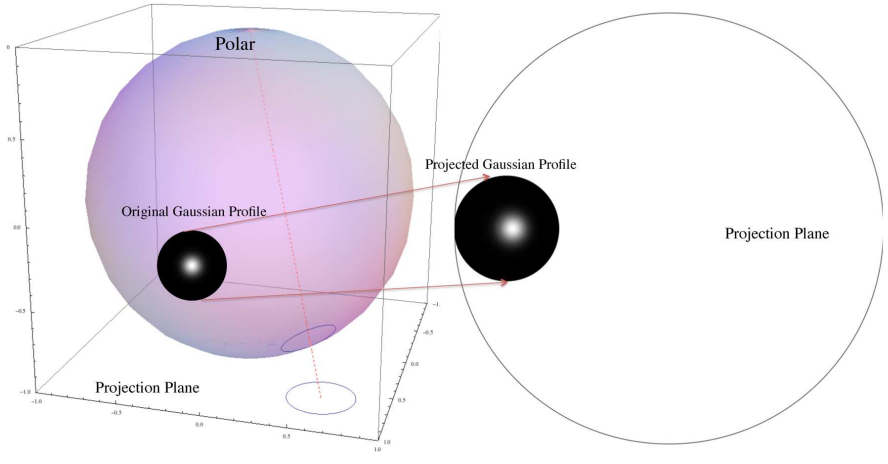

Figure 6.1: STR Projection. A circle on the sphere is projected to a circle on the projection plane.

a singular point at the north pole. To reduce distortion around the polar, we divided the sphere into two hemispheres. The south hemisphere uses the north pole as the projection pivot, and vice versa. The result is shown on (Fig-6.2).

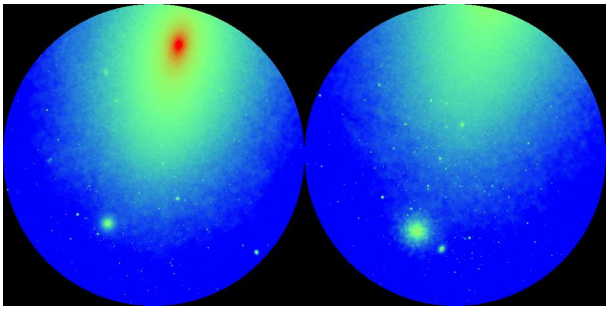

Figure 6.2: Point Sprite Rendering of Annihilation Signal. Every particle's Gaussian profile is projected to a stretched Gaussian. The sphere is divided into two hemispheres to reduce the distortion. Left is the north hemisphere, right is the south hemisphere. Color: Blue: -.5 – Red: 4.5; Units: Log10[GeV^2 cm^{-6} kpc].

6.1.2 Comparasion

Both projection and GPU rendering could introduce additional errors. To verify the result, we converted our result back to the HEALPIX map. As shown in (Fig-6.3), panel (a) the map of the original method, panel (b) the result of the GSP approach. Their difference is shown in panel (c). The bright peaks appear in identical positions as the original method. Some errors are shown on very bright peaks; it may be due to the float point error of GPUs. The largest error is shown on the joint of the North and South sphere. We argue that, with good awareness of these, the physics is not changed and hence the error is negligible [1].

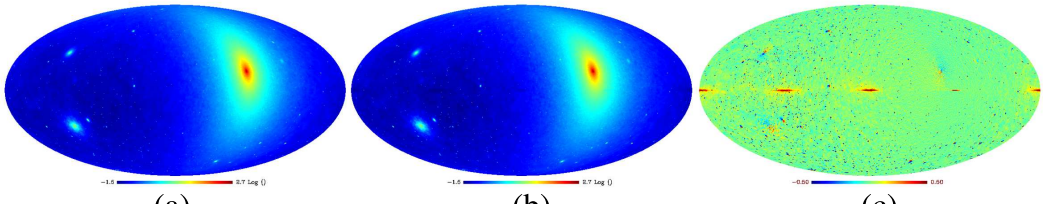

Figure 6.3: Mollweide Projection of the All Sky Flux Maps. (a) is generated by the original slow method; (b) is generated by the GSP method. (c) shows their difference.

Compare to the original HEALPIX (c++/idl version) method (with NSIDE = 512), which runs nearly for 7 hrs to complete the computation on an Intel(R) Xeon(R) CPU X7350 2.93GHz CPU. For the revised CUDA code, running on a NVIDIA GTX 480 GPU, it still costs 1 hr. However, with our method, running on the same GPU, the time consumed is only about 24 seconds (with resolution $512 \times 512 \times 2$ pixels). With an analytical normalization method, the computation time could be reduced to 12 seconds. Moreover, increasing resolution will only lead to a linear increasing on the consuming time, e.g., with resolution

[1] It turns out it is a bug in the code. In the newest version of the code, this discrepancy no longer exists.

$1024 \times 1024 \times 2$ pixels, the time cost for the analytical method is about 20 seconds (all the above measurements do not contain the I/O time).

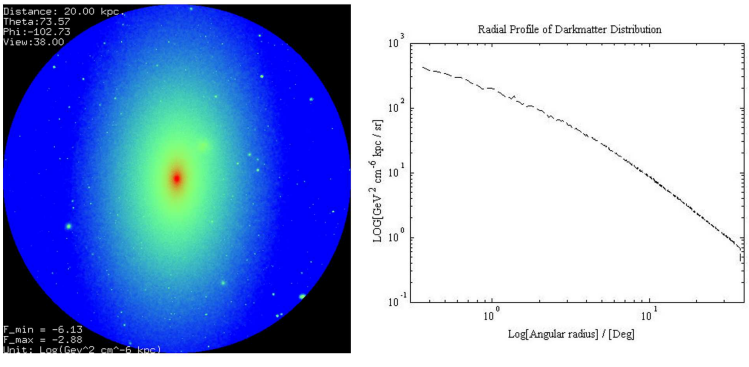

Figure 6.4: Fig-4: GSP Method Works as Virtual Observatory. Left panel simulates observing at the Galaxy center on earth with a view angle of 78 degrees (each pixel is about 9 arcmin, similar to Fermi-LATs observation). Right panel is the radial profile of the signal distribution, restoring Ref.[3]'s Fig-4.

6.1.3 Potential Application

Our method is universal to any dark matter simulation. It could tremendously increase the analysis speed when a simulation is computed out. In addition to generating an all-sky map, it could also work as a virtual observatory. As shown on Fig-6.4 (left panel), by specifying a view angle, and only project a single sphere, we simulate an Gamma-Ray observation (without convolving with a PSF). The resolution used here is about 9 arcmin/pixel, similar to Fermi-LAT's angular resolution. Fig-6.4 (right panel) is angular distribution of the Gamma-ray, restoring Ref.[3]'s Fig-4.

6.1.4 Conclusion

STR projection fits very well to visualize the dark matter annihilation signal. With small error caused by the GPU rendering and re-projection, the speedup obtained by GSP is remarkable. Compared to the previous code, GSP is about 1000 times faster. Our method could be used as a preview of any Dark Matter simulation before being formally analyzed.

6.2 Lagrangian Tessellation Field Estimation

Once a cosmological simulation is carried out, an important task is to estimate the density field of the simulated data. There are many methods of density estimation of an N-body simulation. For instance, the Delaunay and Voronoi Tessellation Field Estimators (DTFE, VTFE),[108] who treat each tracer particle as a physical particle, and assign a volume to each particle using a Voronoi Tessellation. We call a Voronoi tessellation as a Eulerian tessellation, which considers a fixed coordinate system and treats fluid particles as moving objects. In contrast, Lagrangian Tessellation fixes the coordinates system to fluid particles. This is valid in a collisionless fluid since the volume of the phase space is preserved. Thus, for a cosmological simulation with an initial regular grid, we divide each cell, forming by eight tracer particles, into six tetrahedra. See Figure 6.5 as an illustration. Lagrangian Tessellation Field Estimator (LTFE) initially was proposed by Ref.[107,145] In this novel approach, particles define Lagrangian volumes, which deforms under gravity. The deformed volume, in turn, reflects the changes in density field. Fig. 4.1 shows transformed-density slices measured this way, from our simulations. The Lagrangian tessellation has more

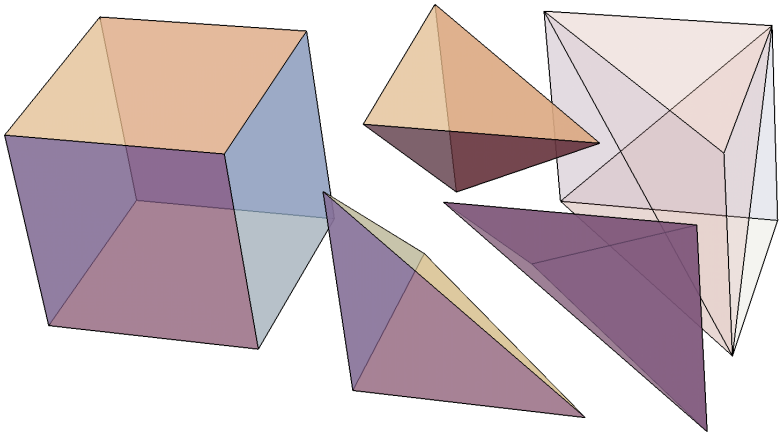

Figure 6.5: A cubic cell (left) is split into 6 tetrahedra.

physical meaning than the Eulerian tessellations. In the LTFE, particles are treated not as mass blobs, but as vertices on a dark-matter sheet (see also[97,146]). The density field is thus formed as if the space is an origami.

Computationally, the difficulty comes from how to get the density on a finer grid. For instance, given a point x in space, to estimate the density $\rho(x)$, we need to sum over all the tetrahedra that take x as an inner point. Each point x may have hundreds of streams crossing it. Thus, for a naïve solution, we might need to search for all the tetrahedra that containing x. Implementation of this kind turns out to be too costly. For instance, it is difficult to implement a query for membership in a tetrahedron. It is also hard to store all tetrahedra in memory to obtain a fast calculation. In any case, suppose we want to estimate

the density of 10^9 positions of a 10^9-particle simulation, the above approach is simply not tractable, i.e. requires $O(10^{18})$ operations.

We provide a streaming algorithm using OpenGL rendering to estimate the density on grid-based positions, e.g., we estimate the densities of 10^9 positions on a 1000^3 uniform grid. To do so, we first convert the original dataset as a set of tetrahedra using the method above. This step takes a single scan over the entire dataset, and the computation time is almost linear in the size of the dataset. Then instead of estimate the 10^9 positions simultaneously, we estimate 1000 sheets of 2D positions, each of which contains 1000×1000 positions. For each sheet, it cut through each tetrahedron at a certain z-coordinate. It turns out that the cut on this sheet is either one triangle, or two triangles (see Figure 6.6). Our algorithm

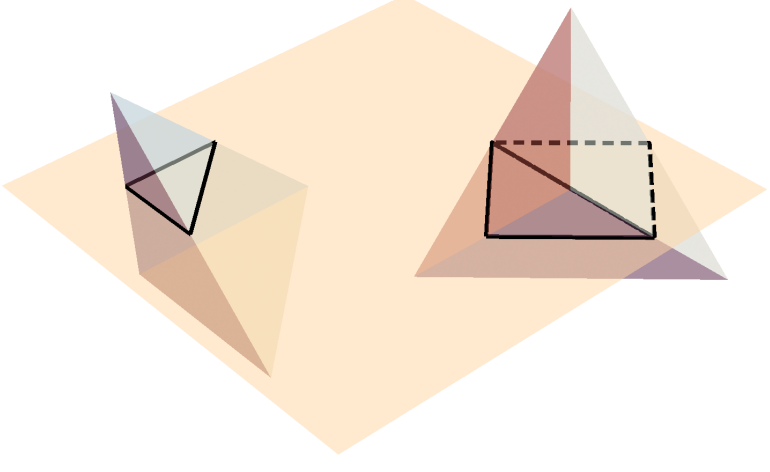

Figure 6.6: A sheet intersects with a tetrahedron. There are either two triangles or one triangle on the sheet.

makes another pass over the tetrahedra data, and convert each tetrahedra into at most 1000 sheets of triangles. This step can be highly parallelized since each tetrahedron, and each sheet can be processed separately. Using CUDA, we can speed the process up to 100 times than a vanilla CPU code. Lastly, for the triangle data, we use OpenGL to render them on the 2D-sheets. Since OpenGL uses the GPU hardware, which is designed for this task, our last step is, in fact, the fastest. Our code can be obtained from Ref.[181]

6.3 Streaming Algorithm Speeds up Halo Finding

This section is heavily based on paper Ref.[182]

Cosmological N-body simulations are essential for studies of the large-scale distribution of matter and galaxies in the Universe. This analysis often involves finding clusters of particles and retrieving their properties. Detecting such "halos" among a very large set of particles is a computationally intensive problem, usually executed on the same supercomputers that produced the simulations, requiring huge amounts of memory.

Recently, a new area of computer science emerged. This area, called streaming algorithms, provides new theoretical methods to compute data analytics in a scalable way using only a single pass over a data sets and logarithmic memory.

The main contribution of this chapter is a novel connection between the analysis of N-body simulations and the streaming algorithms. In particular, we investigate a link between halo finders and the problem of finding frequent items (heavy hitters) in a data

stream, that greatly reduces the computational resource requirements, especially the memory needs. Based on this connection, we build a new halo finder algorithm by running efficient heavy hitter algorithms as a black-box. We implement two representatives of the family of heavy hitter algorithms, the Count-Sketch algorithm (CS) and the Pick-and-Drop sampling (PD), and evaluate their accuracy and memory usage. Comparison with other halo-finding algorithms from[183] shows that our halo finder can locate the largest haloes using significantly smaller memory space and with comparable running time. This streaming approach makes it possible to run and analyze extremely large data sets from N-body simulations on a smaller machine, rather than on supercomputers. Our findings demonstrate the connection between the halo search problem and streaming algorithms as a promising initial direction of further research.

The goal of astrophysics is to explain the observed properties of the universe we live in. In cosmology in particular, one tries to understand how matter is distributed on the largest scales we can observe. In this effort, advanced computer simulations play an ever more important role. Simulations are currently the only way to accurately understand the nonlinear processes that produce cosmic structures such as galaxies and patterns of galaxies. Hence a large amount of effort is spent on running simulations modelling representative parts of the universe in ever greater detail. A necessary step in the analysis of such simulations involves locating mass concentrations, called "haloes", where galaxies would be expected to form. This step is crucial to connect theory to observations – galaxies are the most observable objects that trace the large-scale structure, but their precise spatial distribution is only established through these simulations.

Many algorithms have been developed to find these haloes in simulations. The algorithms vary widely, even conceptually. There is no absolutely agreed-upon physical definition of a halo, although all algorithms give density peaks, i.e. clusters of particles. Galaxies are thought to form at these concentrations of matter. Some codes find regions inside an effective high-density contour, such as Friends-of-Friends (FoF).[184] In FoF, particles closer to each other than a specified linking length are gathered together into haloes. Other algorithms directly incorporate velocity information as well. Another approach finds particles that have crossed each other as compared to the initial conditions, which also ends up giving density peaks.[97] FoF is often considered to be a standard approach, if only because it was among the first used, and is simple conceptually. The drawbacks of FoF include that the simple density estimate can artificially link physically separate haloes together, and the arbitrariness of the linking length. A halo-finding comparison project[185] evaluated the results of 17 different halo-finding algorithms; further analysis appeared in.[183] We take the FoF algorithm as a fiducial result for comparison, but compare to results from some other finders, as well.

Halo-finding algorithms are generally computationally intensive, often requiring all particle positions and velocities to be loaded in memory simultaneously. In fact most are executed during the execution of the simulation itself, requiring comparable computational resources. However, in order to understand the systematic errors in such algorithms, it is often necessary to run multiple halo-finders, often well after the original simulation has been run. Also, many of the newest simulations have several hundred billion to a trillion particles, with a very large memory footprint, making such posterior computations quite

difficult. Here, we investigate a way to apply streaming algorithms as halo finders, and compare the results to those of other algorithms participating in the Halo-Finding Comparison Project.

Recently, streaming algorithms[186] have become a popular way to process massive data sets. In the streaming model, the input is given as a sequence of items and the algorithm is allowed to make a single or constant number of passes over the input data while using sub-linear, usually poly-logarithmic space compared to the storage of the data. Streaming algorithms have found many applications in networking ([187–189]), machine learning ([190,191]), financial analytics ([192–194]) and databases ([195,196]).

In this chapter, we apply streaming algorithms to the area of cosmological simulations and provide space and time efficient solutions to the halo finding problem. In particular, we show a relation between the problem of finding haloes in the simulation data and the well-known problem of finding "heavy hitters" in the streaming data. This connection allows us to employ efficient heavy hitter algorithms, such as Count-Sketch[197] and Pick-and-Drop Sampling.[198] By equating heavy hitters to haloes, we are implicitly defining haloes as positions exceeding some high density threshold. In our case, these usually turn out to be density peaks, but only because of the very spiky nature of the particle distributions in cosmology. Conceptually, FoF haloes are also regions enclosed by high density contours, but in practice, the FoF implementation is very different from ours.

6.3.1 Streaming Algorithm

In this section, we investigate the application of streaming algorithms to find haloes using a strong relation between the halo-finding problem and the heavy hitter problem, which we discuss in section 6.3.2. Heavy hitter algorithms find the k densest regions, that may physically correspond to haloes. In our implementation, we carefully choose k to get the desired outcome. This parameter k is as also discussed in section 6.3.2. We first present in the next sub-section the formal definition of streaming algorithms and the connection between heavy hitter problem and halo-finding problem. After that, we presents the basic procedures of the two heavy hitter algorithms: Count-Sketch and Pick-and-drop Sampling.

6.3.2 Streaming Data Model

Definitions

A data stream $D = D(n, m)$ is an ordered sequence of objects $p_1, p_2, \ldots, p_n$, where $p_j = 1 \ldots m$. The elements of the stream can represent any digital object: integers, real numbers of fixed precisions, edges of a large graphs, messages, images, web pages, etc. In the practical applications both n and m may be very large, and we are interested in the algorithms with $o(n + m)$ space. A streaming algorithm is an algorithm that can make a single pass over the input stream. The above constraints imply that a streaming algorithm is often a randomized algorithm that provides approximate answers with high probability. In practice, these approximate answers often suffice.

We investigate the results of cosmological simulations where the number of particles will soon reach 10^{12}. Compared to offline algorithms that require the input to be entirely in memory, streaming algorithms provide a way to process the data using only megabytes memory instead of gigabytes or terabytes in practice.

Heavy Hitter

For each element i, its frequency f_i is the number of its occurrences in D. The k^{th} frequency moment of a data stream D is defined as $F_k(D) = \sum_{i=1}^{m} f_i^k$. We say that an element is "heavy" if it appears more times than a constant fraction of some L_p norm of the stream, where $L_p = (\sum_i f_i^p)^{1/p}$ for $p > 1$. In this paper, we consider the following heavy hitter problem.

Problem 1. *Given a stream D of n elements, the ϵ-approximate (ϕ, L_p)-heavy hitter problem is to find a set of elements T:*

- $\forall i \in [m],\ f_i > \phi L_p \implies i \in T.$
- $\forall i \in [m],\ f_i < (\phi - \epsilon) L_p \implies i \notin T.$

We allow the heavy hitter algorithms to use randomness; the requirement is that the correct answer should be returned with high probability. The heavy hitter problem is equivalent to the problem of approximately finding the k most frequent elements. Indeed, the top k most frequent elements are in the set of (ϕ, L_1)-heavy hitters in the stream, where $\phi = \Theta(1/k)$. There is a $\Omega(1/\epsilon^2)$ trade-off between the approximation error ϵ and the

memory usage. Heavy hitter algorithms are building blocks of many data stream algorithms ([199, 200]). We treat the cosmological simulation data from[185] as a data stream. To do so, we apply an online transformation that we describe in the next section.

Data Transformation

In a cosmological simulation, dark matter particles form structures through gravitational clustering in a large box with periodic boundary conditions representing a patch of the simulated universe. The box we use[185] is of size 500 Mpc/h, or about 2 billion light-years. The simulation data consists of positions and velocities of 256^3, 512^3 or 1024^3 particles, each representing a huge number of physical dark-matter particles. They are distributed rather uniformly on large scales ($\gtrsim$ 50 Mpc/h) in the simulation box, clumping together on smaller scales. A halo is a clump of particles that are gravitationally bound.

To apply the streaming algorithms, we transform the data. We discretize the spatial coordinates so that we will have a finite number of types in our transformed data stream. We partition the simulation box into a grid of cubic cells, and bin the particles into them. The cell size is chosen to be 1 Mpc/h as to match a typical size of a large halo; there are thus 500^3 cells. This parameter can be modified in practical applications, but it relates to the space and time efficiency of the algorithm. We summarize the data transformation steps as follows.

- Partition the simulation box into grids of cubic cells. Assign each cell a unique integer ID.

- After reading a particle, determine its cell. Insert that cell ID into the data stream.

Using the above transformation, streaming algorithms can process the particles in the same way as an integer data stream.

Heavy Hitter and Dense Cells

For a heavy-hitter algorithm to save memory and time, the distribution of cell counts must be very non-uniform. The simulations begin with an almost uniform lattice of particles, but after gravity clusters them together, the density distribution in cells can be modeled by a lognormal PDF ([137, 201]):

$$P_{LN}^{(1)}(\delta) = \frac{1}{(2\pi\sigma_1^2)^{1/2}} \exp\left\{-\frac{[\ln(1+\delta)+\sigma_1^2/2]^2}{2\sigma_1^2}\right\} \frac{1}{1+\delta}, \tag{6.2}$$

where $\delta = \rho/\bar{\rho} - 1$ is the overdensity, $\sigma_1^2(R) = \ln[1+\sigma_{\rm nl}^2(R)]$, and $\sigma_{\rm nl}^2(R)$ is the variance of the nonlinear density field in spheres of radius R. Our cells are cubic, not spherical; for theoretical estimates, we use a spherical top-hat of the same volume as a cell.

Let N be the number of cells, and P_c be the distribution of the number of particles per cell. The L_p heaviness ϕ_p can be estimated as

$$\phi_p \approx \frac{P_{200}}{(N\langle P_c^p\rangle)^{1/p}}, \tag{6.3}$$

where P_{200} is the number of particles in a cell with density exactly $200\bar{\rho}$. This density threshold is a typical minimum density of a halo, coming from the spherical-collapse model. We theoretically estimated $\sigma_{\rm nl}$ for the cells in our density field by integrating the nonlinear power spectrum (using the fit of,[202] and the cosmological parameters of the simulation) with a spherical tophat window. The grid size in our algorithm is roughly 1.0 Mpc (500^3 cells in total), giving $\sigma_{\rm nl}(\text{Cell}) \approx 10.75$. We estimated $\phi_1 \approx 10^{-6}$ and $\phi_2 \approx 10^{-3}$, matching order-of-magnitude with the measurement of the actual density variance from the simulation cells. These heaviness values are low enough to presume that a heavy-hitter algorithm will efficiently find cells corresponding to haloes.

6.3.3 Streaming Algorithms for Heavy Hitter Problem

The above relation between the halo-finding problem and the heavy hitter problem encourages us to apply efficient streaming algorithms to build a new halo finder. Our halo finder takes a stream of particles, performs the data transformation described in section 6.3.2 and then applies a heavy hitter algorithm to output the approximate top k heavy hitters in the transformed stream. These heavy hitters correspond to the densest cells in the simulation data as described in section 6.3.2. In our first version of the halo finder, we use Count-Sketch algorithm[197] and Pick-and-Drop Sampling.[198]

The Count-Sketch Algorithm

For a more generalized description of the algorithm, please refer to.[197] For completeness, we summarize the algorithm as follows. The Count-Sketch algorithm uses a compact data structure to maintain the approximate counts of the top k most frequent elements in a stream. This data structure is an $r \times t$ matrix M representing estimated counts for all elements. These counts are calculated by two sets of hash functions: let $h_1, h_2, \ldots, h_r$ be r hash functions, mapping the input items to $\{1, \ldots, t\}$, where each h_i is sampled uniformly from the hash function set H. Let $s_1, s_2, \ldots, s_r$ be hash functions, mapping the input items to $\{+1, -1\}$, uniformly sampled from another hash function set S. We can interpret this matrix as an array of r hash tables, each containing t buckets.

There are two operations on the Count-Sketch data structure. Denote $M_{i,j}$ as the j^{th} bucket in the i^{th} hash table:

- $Add(M, p)$: For $i \in [1, r]$, $M_{i,h_i[p]} += s_i[p]$.
- $Estimate(M, p)$, return $median_i\{h_i[p] \cdot s_i[p]\}$

The *Add* operation updates the approximate frequency for each incoming element and the *Estimate* operation outputs the current approximate frequency. To maintain and store the estimated k most frequent elements, CountSketch also needs a priority queue data structure. The pseudocode of Count-Sketch algorithm is presented in Figure 6.7. More details and theoretical guarantees are presented in.[197]

```
1: procedure CountSketch(r, t, k, D)                    ▷ D is a stream
2:     Initialize an empty r × t matrix M.
3:     Initialize an min-priority queue Q of size k
4:     (particle with smallest count is on the top).
5:     for i = 1, ..., n and p_i ∈ D do
6:         Add(M, p_i);
7:         if p_i ∈ Q then
8:             P_i.count++;
9:         else if Estimate(M, p_i) > Q.top().count then
10:            Q.pop();
11:            Q.push(p_i);
12:        end if
13:    end for
14:    return Q
15: end procedure
```

Figure 6.7: Count-Sketch Algorithm

The Pick-and-Drop Sampling Algorithm

Pick-and-Drop Sampling is a sampling-based streaming algorithm to approximate the heavy hitters. To describe the idea of Pick-and-Drop sampling, we view the data stream as a sequence of r blocks of size t. Define $d_{i,j}$ as the j^{th} element in the i^{th} block and $d_{i,j} = p_{k(i-1)+j}$ in stream D. In each block of the stream, Pick-and-Drop sampling will pick one random sample and record its remaining frequency in the block. The algorithm maintains a sample with the largest current counter and drops previous samples. The pseudocode of Pick-and-Drop sampling[198] is given in Figure 6.8 and we need the following definitions in Figure 6.8. For $i \in [r]$, $j, s \in [t]$, $q \in [m]$ define:

$$f_{i,q} = |\{j \in [t] : d_{i,j} = q\}|, \tag{6.4}$$

$$a_{i,s} = |\{j^* : s \leq j^* \leq t, d_{i,j^*} = d_{i,s}\}|. \tag{6.5}$$

1: **procedure** PICKDROP(r, t, λ, D)
2: Sample S_1 uniformly at random on $[t]$.
3: $L_1 \leftarrow d_{1,S_1}$,
4: $C_1 \leftarrow a_{1,S_1}$,
5: $u_1 \leftarrow 1$.
6: **for** $i = 2, \ldots, r$ **do**
7: Sample S_i uniformly at random on $[t]$.
8: $l_i \leftarrow d_{i,S_i}$, $c_i \leftarrow a_{i,S_i}$
9: **if** $C_{i-1} < \max(c_i, \lambda u_{i-1})$ **then**
10: $L_i \leftarrow l_i$,
11: $C_i \leftarrow c_i$,
12: $u_i \leftarrow 1$
13: **else**
14: $L_i \leftarrow L_{i-1}$,
15: $C_i \leftarrow C_{i-1} + f_{i,L_{i-1}}$,
16: $u_i \leftarrow q_{i-1} + 1$
17: **end if**
18: **end for**
19: **return** $\{L_r, C_r\}$
20: **end procedure**

Figure 6.8: Pick-and-Drop Algorithm

The detail implementation is in Section 6.3.4.

6.3.4 Implementation

Simulation Data

The N-body simulation data we use as the input to our halo finder was used in the halo-finding comparison project[185] and consists of various resolutions (numbers of particles) of the MareNostrum Universe cosmological simulation.[203] These simulations ran in a 500 h^{-1}Mpc box, assuming a standard ΛCDM (cold dark matter and cosmological constant) cosmological model.

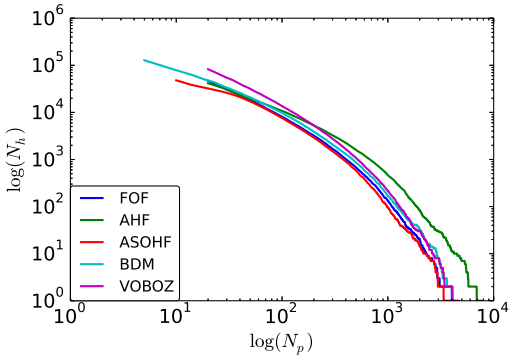

Figure 6.9: Halo mass distribution of various halo finders.

In the first implementation of our halo finder, we consider two halo properties: center position and mass (the number of particles in it). We compare to the the fiducial offline algorithm FoF. The distributions of halo sizes from different halo finders are presented in Fig. 6.9.

Since our halo finder builds on the streaming algorithms of finding frequent items, the algorithms need to transform the data as described in section 6.3.2 — dividing all the particles into different small cells and label each particle with its associated cell ID. For example, if an input dataset contains three particles p_1, p_2, p_3 and they are all included in a cell of ID $= 1$, then the transformed data stream becomes $1, 1, 1$. The most frequent element in the stream is obviously 1 and thus the cell 1 is the heaviest cell overall.

Implementation Details

Our halo finder implementation is written using C++ with GNU GCC compiler 4.9.2. We implemented Count-Sketch and Pick-and-Drop sampling as two algorithms to find heavy hitters.

Count-Sketch-based Halo Finder There are three basic steps in the Count-Sketch algorithm, which returns the heavy cells and the number of particles associated with them. (1) Allocate memory for the CountSketch data structure to hold current estimates of cell frequencies; (2) use a priority queue to record the k most frequent elements; (3) return the positions of the top k heavy cells. Figure 6.10 presents the process of the Count-Sketch.

The Count-Sketch data structure is an $r \times t$ matrix. Following,[204] we set $r = \log(\frac{n}{\epsilon})$ and t to be sufficiently large (>1 million) to achieve an expected approximation error $\epsilon = 0.05$. We build the matrix as a 2D array with $r \times t$ 0's. For each incoming element in the stream, an *Add* operation has to be executed and an *estimate* operation needs to be executed only when this element is not in the queue.

Pick-and-Drop-based Halo Finder

In the Pick-and-Drop sampling based halo finder, we implement a general hash function H: $\mathbb{N}^+ \rightarrow \{1, 2, \ldots, ck\}$, where $c \geq 1$, to gain the probability of success to approximate

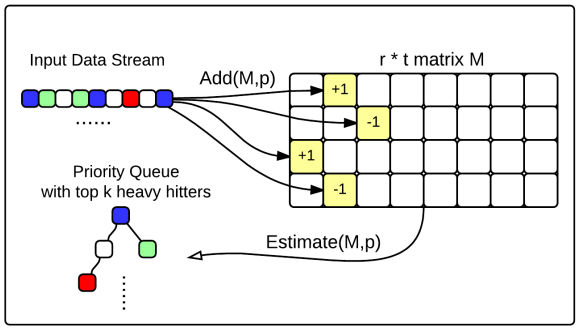

Figure 6.10: Count-Sketch Algorithm

the k heaviest cells. We apply the hash function H on every incoming element and put the elements with the same hash value together such that the original stream is divided into ck smaller sub-streams. Meanwhile, we initialize ck instances of Pick-and-Drop sampling so that each PD instance will process one sub-stream. The whole process of approximating the heavy hitters is presented in Figure 6.11. In this way, the repeated items in the whole stream will be distributed into the same sub-stream and they are much heavier in this sub-stream. With high probability, each instance of Pick-and-Drop sampling will output the heaviest one in each of the sub-streams, and in total we will have ck output items. Because of the randomness in the sampling method, we will expect some of inaccurate heavy hitters among the total ck outputs. By setting a large c, most of the actual top k most frequent elements should be inside the ck outputs (raw result).

To get precise properties of haloes, such as the center, and mass, an offline algorithm such as FoF[184] can be applied to the particles inside the returned heavy cells and their neighbor cells. This needs an additional pass over the data but we only need to store a

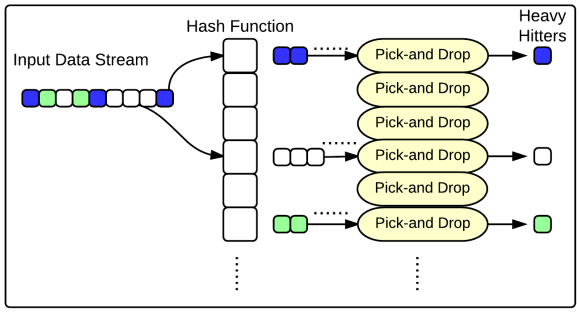

Figure 6.11: Pick-and-Drop Sampling

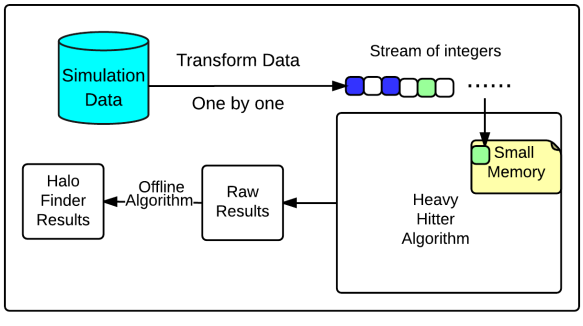

Figure 6.12: Halo Finder Procedure

small amount of particles to run those offline in-memory algorithms. The whole process of the halo finder is represented in Figure 6.12, where heavy hitter algorithms can be regarded as a black box. That is, any theoretically efficient heavy hitter algorithms could be applied to further improve the memory usage and practical performance.

Shifting Method

In the first pass of our halo finder, we only use the position of a heavy cell as the position of a halo. However, each heavy cell may contain several haloes and some of the haloes located on the edges between two cells cannot be recognized because the cell size in the data transformation step is fixed. To recover those missing haloes, we utilize a simple shifting method:

- Initialize 2^d instances of Count-Sketch or Pick-and-Drop in parallel, where d is the dimension. Our simulation data reside in three dimensions, so $d = 3$.

- Move all the particles to one of the 2^d directions with a distance of 0.5 Mpc/h (half of the cell size). In each of the 2^d shifting processes, assign a Count-Sketch/Pick-and-Drop instance to run. By combining the results from 2^d shifting processes, we expect that the majority of k largest haloes are discovered. All the parallel instances of the CountSketch/Pick-and-Drop are enabled by OpenMP 4.0 in C++.

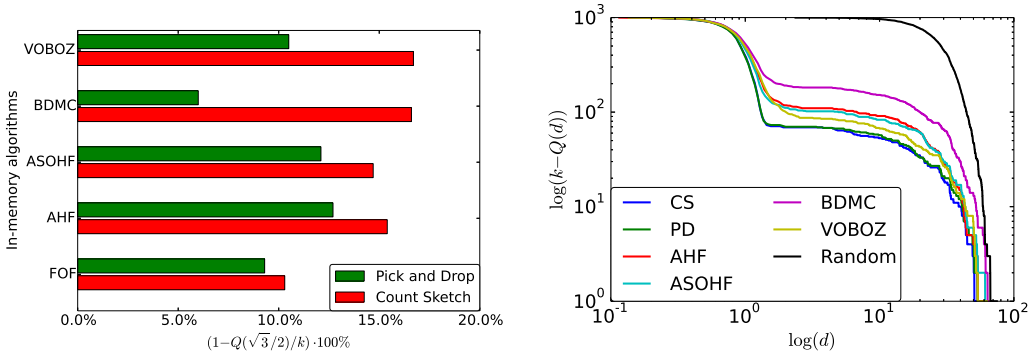

Figure 6.13: Left Panel: Measures of the disagreement between PD and CS, and various in-memory algorithms. The percentage shown is the fraction of haloes farther than a half-cell diagonal ($0.5\sqrt{3}$ Mpc/h) from PD or CS halo positions. Right Panel: The number of top-1000 FoF haloes farther than a distance d away from any top-1000 halo from the algorithm of each curve.

6.3.5 Evaluation

To evaluate how well streaming based halo finders work, we mainly focus on testing it in the following four aspects:

- **Correctness:** Evaluate how close are the positions of k largest haloes found by the streaming-based algorithms to the top k large haloes returned by some widely used in-memory algorithms. Evaluate the trade-off between the selection k and the quality of result.

- **Stability:** Since streaming algorithms always require some randomness and may produce some incorrect results, we want to see how stable are streaming based heavy hitter algorithms are.

- **Memory Usage:** Linear memory space requirement is a "bottle neck" for all offline algorithms, and it is the central problem that we are trying to overcome by applying

streaming approach. Thus it is significantly important to theoretically or experimentally estimate the memory usage of Pick-and-Drop and Cound-sketch algorithms.

In the evaluation, all the in-memory algorithms we choose to compare were proposed in the Halo-Finding Comparison Project.[185] We test against the fiducial FOF method, as well as four others that find density peak:

1. **FOF** by Davis et al.[184]

   "Plain-vanilla" Friends-of-Friends.

2. **AHF** by Knollmann & Knebe[205]

   Density peaks search with recursively refined grid

3. **ASOHF** by Planelles & Quilis.[206]

   Finds spherical-overdensity peaks using adaptive density refinement.

4. **BDM**,[207] run by Klypin & Ceverino "Bound Density Maxima" – finds gravitationally-bound spherical-overdensity peaks.

5. **VOBOZ** by Neyrinck et al[208]

   "Voronoi BOund Zones" – finds gravitationally bound peaks using a Voronoi tessellation.

Correctness

As there is no agreed upon rule how to define the center and the boundary of a halo, it is impossible to theoretically define and deterministically verify the correctness of any halo finder. Therefore a comparison to the results of previous widely accepted halo finders seems to be the best practical verification of a new halo finder. To compare the outputs of two different halo finders we need to introduce some formal measure of similarity. The most straight forward way to compare them is to consider one of them H as a ground truth, and another one E as an estimator. Among this the FOF algorithm is considered to be the oldest and the most widely used, thus in our initial evaluation we decided to concentrate on the comparison with FOF. Then the most natural measure of similarity is number of elements in H that match to elements in output of E. More formally we will define "matches" as: for a given θ we will say that center $e_i \in E$ matches the element $h_i \in E$ if $dist(e_i, h_i) \leq \theta$, where $dist(\cdot, \cdot)$ is Euclidean distance. Then our measure of similarity is:

$$Q(\theta) = Q(E_k, H_k, \theta) = |\{h_i \in H_k : \min_{e_j \in E_k} dist(h_i, e_j) < \theta\}|,$$

where k represents k heaviest halos.

We compare the output of both streaming-based halo finders to the output of in-memory halo finders. We made comparisons for the $256^3, 512^3$ and 1024^3-particle simulations, finding the top 1000 and top 10000 heavy hitters. Since the comparison results in all cases were similar, the figures presented below are for the 256^3 dataset, and $k = 1000$.

On the Figure 6.13 we show for each in-memory algorithm the percentage of centers

that were not found by streaming-based halo finder. We can see that both the Count-Sketch and Pick-and-Drop algorithms missed not much more than 10 percent of the haloes in any of the results from the in-memory algorithms.

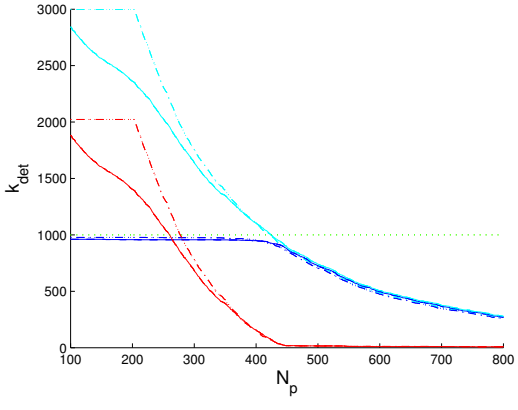

Figure 6.14: Number of detected halos by our two algorithms. The solid lines correspond to (CS) and the dashed lines to (PD). The dotted line at $k = 1000$ shows our selection criteria. The x axis is the threshold in the number of particles allocated to the heavy hitter. The cyan color denotes the total number of detections, the blue curves are the true positives (TP), and the red curves are ethe false positives (FP).

To understand whether the 10 percent means two halo catalogs are close to each other or not, we will choose one of the in-memory algorithms as a ground truth and compare how close the other in-memory algorithms are. Again, we choose FOF algorithm as a ground truth. The comparison is depicted in Fig. 6.13. From this graph you can see that the outputs of Count-Sketch and Pick-and-Drop based halo finders are closer to the FOF haloes, than other in-memory algorithms. It can be easily explained, as after finding heavy cells we apply the same FOF to these heavy cells and their neighborhoods, the output should always have similar structure to the output of in-memory FOF on the full dataset. Also from this graph you will see that each line can be represented as a mix of two components, one of which is the component of random distribution. It means that after a distance of

$\sqrt{3}/2$ all matches are the same if we just put bunch of points at random.

The classifier is using a top-k to select the halo candidates. Figure 6.14 shows how sensitive the results are to the selection threshold of $k = 1000$. It shows several curves, including the total number of heavy hitters, the ones close to an FoF group – we can call these true positive (TP) – and the ones detected, but not near an FoF object (false positives FP). From the figure, it is clear that the threshold of 1000 is close to the optimal detection threshold, preserving TP and minimizing FP. This corresponds to a true positive detection rate (TPR) of 96% and a false positive detection rate of 3.6%. If we lowered our threshold to $k = 900$, our TPR drops to 91% but the FPR becomes even lower, 0.88%.

These tradeoffs can be shown on a so-called ROC-curve (receiver operating characteristic), where the TPR is plotted against the FPR. This shows how lowering the detection threshold increases the true detections, but the false detection rate increases much faster. Using the ROC curve, shown below we can see the position of the $k = 1000$ threshold as a circle and the $k = 900$ as a square.

Finally, we should also ask, besides the set comparison, how do the individual particle cardinalities counted around the heavy hitters correspond to the FoF ones. Our particle counting is restricted to neighboring cells, while the FoF is not, so we will always be undercounting. To be less sensitive to such biases, we compare the rank ordering of the two particle counts in the two samples in Fig. 6.16. The rank 1 is assigned to the most massive objects in each set.

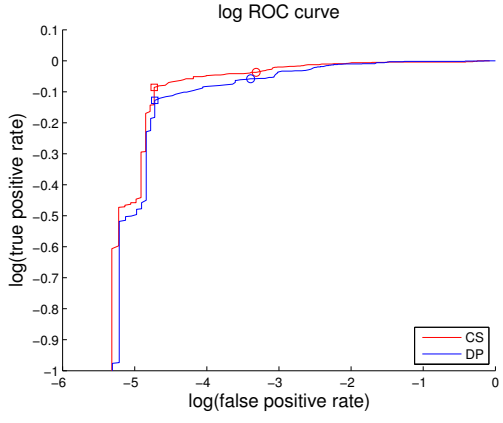

Figure 6.15: This ROC curve shows the tradeoff between true and false detections as a function of threshold. The figure plots TPR vs FPR on a log-log scale. The two thresholds are shown with symbols, the circle denotes 1000, and the square is 900.

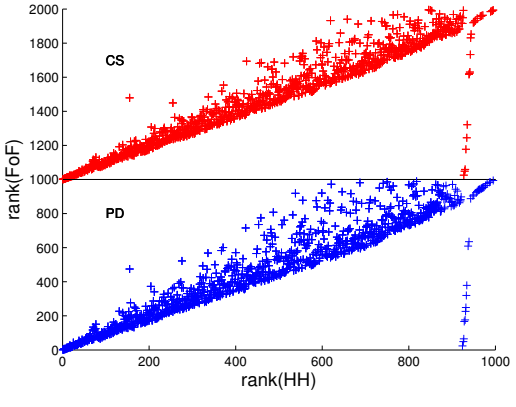

Figure 6.16: The top 1000 heavy hitters are rank-ordered by the number of their particles. We also computed a rank of the corresponding FoF halo. The linked pairs of ranks are plotted. One can see that if we adopted a cut at $k = 900$, it would eliminate a lot of the false positives.

Stability

As most of the streaming algorithms utilize randomness, we estimate how stable our results are compared to the results from a deterministic search. In the deterministic search algorithm, we find the actual heavy cells by counting the number of particles inside them;

we perform the comparison for the dataset containing 256^3 particles. To perform this evaluation we run 50 instances of each algorithm (denoting the outputs as $\{C_{cs}^i\}_{i=1}^{50}$ and $\{C_{pd}^i\}_{i=1}^{50}$). We also count the number of cells of each result that match the densest cells returned by the deterministic search algorithm C_{ds}. The normalized number of matches will be $\rho_{pd}^i = \frac{|C_{pd}^i \cap C_{ds}|}{|C_{ds}|}$ and $\rho_{cs}^i = \frac{|C_{cs}^i \cap C_{ds}|}{|C_{ds}|}$ correspondingly. Our experiment showed:

$$\mu(\rho_{cs}^i) = 0.946,\ \sigma(\rho_{cs}^i) = 2.7 \cdot 10^{-7}$$

$$\mu(\rho_{pd}^i) = 0.995,\ \sigma(\rho_{pd}^i) = 6 \cdot 10^{-7}$$

This means that the approximation error caused by randomness is very small compared with the error caused by transition from overdense cells to halo centers. This fact can also be caught from the Fig. 6.17. On that figure you can see that shaded area below and above the red line and green line, which represents the range of outputs among 50 instances, is very thin. Thus the output is very stable.

Memory Usage

Comparing with current halo finding solutions, streaming approachs' low memory usage is one of the most significant advantages. To the best of our knowledge even for the problem of locating 1000 largest haloes in the simulation data with 1024^3 particles, there is no way to run other halo finding algorithms on a regular PC since 1024^3 particles already need $\approx$ 12GB memory to only store all the particle coordinates; a computing cluster or even supercomputer is necessary. Therefore, the application of streaming techniques introduces

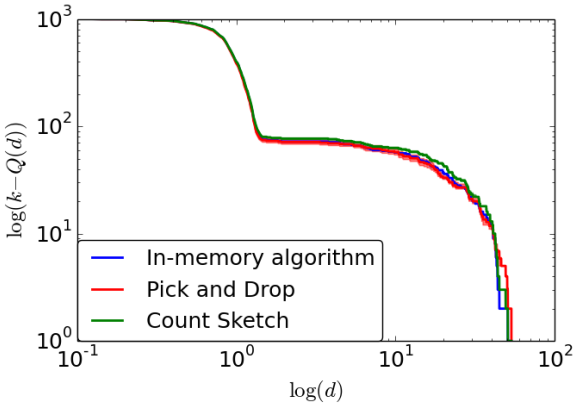

Figure 6.17: Each line on the graph represents the top 1000 halo centers found with Pick-and-Drop sampling, Count-Sketch, and in-memory algorithms, as described in section 6.3.4. The comparison with FOF is shown in Fig.6.13. The shaded area (too small to be visible) shows the variation due to randomness.

a new direction on the development of halo-finding algorithms.

To find top k heavy cells, Count-Sketch theoretically requires following amount of space:

$$O(k \log \frac{n}{\delta} + \frac{\sum_{q'=k+1}^{m} f_{q'}^2}{(\epsilon f_k)^2} \log \frac{n}{\delta}),$$

where $1 - \delta$ is probability of success, ϵ is an Q_k estimation error, and Q_k is the frequency of k-th heaviest cell. It is worth mentioning that in application to the heavy cell searching problem the second term is the dominating one. The first factor in the second term represents the linear dependency of memory usage on the heaviness of top k cells. Thus we can expect linear memory usage for small dataset. But as dataset grows the dependency becomes logarithmic if we assume the same level of heaviness. Experiments verify this observation, as for small dataset with 256^3 particles Count-Sketch algorithm used around 900 megabytes memory, while for the large 1024^3-dataset, the memory usage was increased to

nearly 1000 megabytes. Thus the memory grows logarithmically with the size of dataset if we assume almost constant heaviness of the top k cells; that is why such approach is scalable for even larger datasets.

In the experiments using this particular simulation data, Pick-and-Drop sampling shows much better performance in terms of memory usage than Count-Sketch. The actual usage of memory was around 20 megabytes for the dataset with 256^3 particles and around 30 megabytes for the dataset with 1024^3 particles.

6.3.6 Conclusion

In this paper we find a novel connection between the problem of finding the most massive halos in cosmological N-Body simulations and the problem of finding heavy hitters in data streams. According to this link, we have built a halo finder based on the implementation of Count-Sketch algorithm and Pick-and-Drop sampling. The halo finder successfully locates most ($> 90\%$) of the k largest haloes using sub-linear memory. Most halo-finders require the entire simulation to be loaded into memory. But our halo finder does not and could be run on the massive N-body simulations that are anticipated to arrive in the near future with relatively modest computing resources. We will continue to improve the performance of our halo finder, something we have as yet not paid much attention to. In the very first implementation we evaluated here, we mainly focus on the verification of precision instead of performance. But both Count-Sketch and Pick-and-Drop sampling can be easily parallelized further to achieve significantly better performance. The majority of the computation on Count-Sketch is spent on the calculations of $r \times t$ hash functions. A straight forward

way to improve the performance is taking advantage of the highly parallel GPU streaming processors to improve the performance of calculating a large number of hash functions. Similarly, Pick-and-Drop sampling is also a good candidate for more parallelism since the Pick-and-Drop instances are running independently.

We also note that this halo finder finds only the k most massive haloes. These are features of interest in the simulation, but some further work is required for our methods to return a complete set of haloes as an in-memory algorithm.

Bibliography

[1] J. Diemand, M. Kuhlen, P. Madau, M. Zemp, B. Moore, D. Potter, and J. Stadel, "Clumps and streams in the local dark matter distribution," *Nature*, vol. 454, no. 7205, pp. 735–738, 2008.

[2] M. C. Neyrinck, "Removable Matter-power-spectrum Covariance from Bias Fluctuations," *The Astrophysical Journal*, vol. 736, p. 8, Jul. 2011.

[3] M. Kuhlen, J. Diemand, and P. Madau, "The dark matter annihilation signal from galactic substructure: predictions for glast," *The Astrophysical Journal*, vol. 686, no. 1, p. 262, 2008.

[4] F. Zwicky, "Die rotverschiebung von extragalaktischen nebeln," *Helvetica Physica Acta*, vol. 6, pp. 110–127, 1933.

[5] T. Van Albada, K. Begeman, R. Sanscisi, and J. Bahcall, "Distribution of dark matter in the spiral galaxy ngc 3198," *Dark Matter in the Universe*, vol. 4, p. 7, 2004.

[6] P. Ade, N. Aghanim, M. Arnaud, M. Ashdown, J. Aumont, C. Baccigalupi, A. Banday,

R. Barreiro, N. Bartolo, E. Battaner *et al.*, "Planck 2015 results-xiv. dark energy and modified gravity," *Astronomy & Astrophysics*, vol. 594, p. A14, 2016.

[7] K. Griest, "The nature of the dark matter," *arXiv preprint astro-ph/9510089*, 1995.

[8] U. Briel, J. Henry, and H. Böhringer, "Observation of the coma cluster of galaxies with rosat during the all-sky survey," *Astronomy and Astrophysics*, vol. 259, pp. L31–L34, 1992.

[9] P. Fischer and J. A. Tyson, "The mass distribution of the most luminous x-ray cluster rxj1347. 5-1145 from gravitational lensing," *arXiv preprint astro-ph/9703189*, 1997.

[10] D. Clowe, M. Brada**c**, A. H. Gonzalez, M. Markevitch, S. W. Randall, C. Jones, and D. Zaritsky, "A direct empirical proof of the existence of dark matter," *The Astrophysical Journal Letters*, vol. 648, no. 2, p. L109, 2006.

[11] D. J. Eisenstein, D. H. Weinberg, E. Agol, H. Aihara, C. A. Prieto, S. F. Anderson, J. A. Arns, É. Aubourg, S. Bailey, E. Balbinot *et al.*, "Sdss-iii: Massive spectroscopic surveys of the distant universe, the milky way, and extra-solar planetary systems," *The Astronomical Journal*, vol. 142, no. 3, p. 72, 2011.

[12] G. Jungman, M. Kamionkowski, and K. Griest, "Supersymmetric dark matter," *Physics Reports*, vol. 267, no. 5-6, pp. 195–373, 1996.

[13] C. S. Frenk and S. D. White, "Dark matter and cosmic structure," *Annalen der Physik*, vol. 524, no. 9-10, pp. 507–534, 2012.

[14] M. Milgrom and R. H. Sanders, "Modified newtonian dynamics and the dearth of

dark matter in ordinary elliptical galaxies," *The Astrophysical Journal Letters*, vol. 599, no. 1, p. L25, 2003.

[15] Z. Ahmed, D. Akerib, S. Arrenberg, M. Attisha, C. Bailey, L. Baudis, D. Bauer, J. Beaty, P. Brink, T. Bruch *et al.*, "Search for weakly interacting massive particles with the first five-tower data from the cryogenic dark matter search at the soudan underground laboratory," *Physical Review Letters*, vol. 102, no. 1, p. 011301, 2009.

[16] J. Angle, E. Aprile, F. Arneodo, L. Baudis, A. Bernstein, A. Bolozdynya, L. Coelho, C. Dahl, L. Deviveiros, A. Ferella *et al.*, "Constraints on inelastic dark matter from xenon10," *Physical Review D*, vol. 80, no. 11, p. 115005, 2009.

[17] M. Alfonsi, "The xenon dark matter project: from xenon100 to xenon1t," *Nuclear and Particle Physics Proceedings*, vol. 273, pp. 373 – 377, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2405601415005428

[18] D. S. Akerib, H. M. Araújo, X. Bai, A. J. Bailey, J. Balajthy, S. Bedikian, E. Bernard, A. Bernstein, A. Bolozdynya, A. Bradley, D. Byram, S. B. Cahn, M. C. Carmona-Benitez, C. Chan, J. J. Chapman, A. A. Chiller, C. Chiller, K. Clark, T. Coffey, A. Currie, A. Curioni, S. Dazeley, L. de Viveiros, A. Dobi, J. Dobson, E. M. Dragowsky, E. Druszkiewicz, B. Edwards, C. H. Faham, S. Fiorucci, C. Flores, R. J. Gaitskell, V. M. Gehman, C. Ghag, K. R. Gibson, M. G. D. Gilchriese, C. Hall, M. Hanhardt, S. A. Hertel, M. Horn, D. Q. Huang, M. Ihm, R. G. Jacobsen, L. Kastens, K. Kazkaz, R. Knoche, S. Kyre, R. Lander, N. A. Larsen, C. Lee, D. S. Leonard, K. T. Lesko, A. Lindote, M. I. Lopes, A. Lyashenko, D. C. Malling,

R. Mannino, D. N. McKinsey, D.-M. Mei, J. Mock, M. Moongweluwan, J. Morad, M. Morii, A. S. J. Murphy, C. Nehrkorn, H. Nelson, F. Neves, J. A. Nikkel, R. A. Ott, M. Pangilinan, P. D. Parker, E. K. Pease, K. Pech, P. Phelps, L. Reichhart, T. Shutt, C. Silva, W. Skulski, C. J. Sofka, V. N. Solovov, P. Sorensen, T. Stiegler, K. O'Sullivan, T. J. Sumner, R. Svoboda, M. Sweany, M. Szydagis, D. Taylor, B. Tennyson, D. R. Tiedt, M. Tripathi, S. Uvarov, J. R. Verbus, N. Walsh, R. Webb, J. T. White, D. White, M. S. Witherell, M. Wlasenko, F. L. H. Wolfs, M. Woods, and C. Zhang, "First results from the lux dark matter experiment at the sanford underground research facility," *Phys. Rev. Lett.*, vol. 112, p. 091303, Mar 2014. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevLett.112.091303

[19] N. Kaiser and G. Squires, "Mapping the dark matter with weak gravitational lensing," *The Astrophysical Journal*, vol. 404, pp. 441–450, 1993.

[20] O. Adriani, G. Barbarino, G. Bazilevskaya, R. Bellotti, M. Boezio, E. Bogomolov, L. Bonechi, M. Bongi, V. Bonvicini, S. Borisov *et al.*, "Measurements of quasi-trapped electron and positron fluxes with pamela," *Journal of Geophysical Research: Space Physics*, vol. 114, no. A12, 2009.

[21] O. Adriani, G. Barbarino, G. Bazilevskaya, R. Bellotti, M. Boezio, E. Bogomolov, M. Bongi, V. Bonvicini, S. Borisov, S. Bottai *et al.*, "Cosmic-ray electron flux measured by the pamela experiment between 1 and 625 gev," *Physical Review Letters*, vol. 106, no. 20, p. 201101, 2011.

[22] A. Panov, J. Adams, H. Ahn, G. Bashindzhagyan, K. Batkov, J. Chang, M. Christl,

A. Fazely, O. Ganel, R. Gunasingha *et al.*, "The energy spectra of heavy nuclei measured by the atic experiment," *Advances in Space Research*, vol. 37, no. 10, pp. 1944–1949, 2006.

[23] M. Ackermann, M. Ajello, W. Atwood, L. Baldini, J. Ballet, G. Barbiellini, D. Bastieri, K. Bechtol, R. Bellazzini, B. Berenji *et al.*, "Fermi-lat observations of the diffuse γ-ray emission: implications for cosmic rays and the interstellar medium," *The Astrophysical Journal*, vol. 750, no. 1, p. 3, 2012.

[24] M. Ackermann, M. Ajello, A. Allafort, W. Atwood, L. Baldini, G. Barbiellini, D. Bastieri, K. Bechtol, R. Bellazzini, B. Berenji *et al.*, "Measurement of separate cosmic-ray electron and positron spectra with the fermi large area telescope," *Physical Review Letters*, vol. 108, no. 1, p. 011103, 2012.

[25] F. Aharonian, A. Akhperjanian, A. Bazer-Bachi, M. Beilicke, W. Benbow, D. Berge, K. Bernlöhr, C. Boisson, O. Bolz, V. Borrel *et al.*, "Observations of the sagittarius dwarf galaxy by the hess experiment and search for a dark matter signal," *Astroparticle Physics*, vol. 29, no. 1, pp. 55–62, 2008.

[26] W. H. Press and P. Schechter, "Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation," *The Astrophysical Journal*, vol. 187, pp. 425–438, Feb. 1974.

[27] M. Boylan-Kolchin, V. Springel, S. D. White, A. Jenkins, and G. Lemson, "Resolving cosmic structure formation with the millennium-ii simulation," *Monthly Notices of the Royal Astronomical Society*, vol. 398, no. 3, pp. 1150–1164, 2009.

[28] V. Springel, S. D. White, A. Jenkins, C. S. Frenk, N. Yoshida, L. Gao, J. Navarro, R. Thacker, D. Croton, J. Helly *et al.*, "Simulations of the formation, evolution and clustering of galaxies and quasars," *nature*, vol. 435, no. 7042, pp. 629–636, 2005.

[29] R. Angulo, V. Springel, S. White, A. Jenkins, C. Baugh, and C. Frenk, "Scaling relations for galaxy clusters in the millennium-xxl simulation," *Monthly Notices of the Royal Astronomical Society*, vol. 426, no. 3, pp. 2046–2062, 2012.

[30] J. Kim, C. Park, G. Rossi, S. M. Lee, and J. R. Gott III, "The new horizon run cosmological n-body simulations," *arXiv preprint arXiv:1112.1754*, 2011.

[31] J.-M. Alimi, V. Bouillot, Y. Rasera, V. Reverdy, P.-S. Corasaniti, I. Balmes, S. Requena, X. Delaruelle, and J.-N. Richet, "First-ever full observable universe simulation," in *High Performance Computing, Networking, Storage and Analysis (SC), 2012 International Conference for*. IEEE, 2012, pp. 1–11.

[32] V. Springel, J. Wang, M. Vogelsberger, A. Ludlow, A. Jenkins, A. Helmi, J. F. Navarro, C. S. Frenk, and S. D. White, "The aquarius project: the subhaloes of galactic haloes," *Monthly Notices of the Royal Astronomical Society*, vol. 391, no. 4, pp. 1685–1711, 2008.

[33] J. Stadel, D. Potter, B. Moore, J. Diemand, P. Madau, M. Zemp, M. Kuhlen, and V. Quilis, "Quantifying the heart of darkness with ghalo–a multibillion particle simulation of a galactic halo," *Monthly Notices of the Royal Astronomical Society: Letters*, vol. 398, no. 1, pp. L21–L25, 2009.

[34] L. F. Yang, J. Silk, A. S. Szalay, R. F. Wyse, B. Bozek, and P. Madau, "Dark matter contribution to galactic diffuse gamma ray emission," *Physical Review D*, vol. 89, no. 6, p. 063530, 2014.

[35] H. An, M. Pospelov, J. Pradler, and A. Ritz, "Direct detection constraints on dark photon dark matter," *Physics Letters B*, vol. 747, pp. 331–338, 2015.

[36] M. Aguilar, D. Aisa, B. Alpat, A. Alvino, G. Ambrosi, K. Andeen, L. Arruda, N. Attig, P. Azzarello, A. Bachlechner *et al.*, "Precision measurement of the (e++ e-) flux in primary cosmic rays from 0.5 gev to 1 tev with the alpha magnetic spectrometer on the international space station," *Physical review letters*, vol. 113, no. 22, p. 221102, 2014.

[37] L. Bergström, T. Bringmann, I. Cholis, D. Hooper, and C. Weniger, "New limits on dark matter annihilation from alpha magnetic spectrometer cosmic ray positron data," *Physical review letters*, vol. 111, no. 17, p. 171101, 2013.

[38] M. Madhavacheril, N. Sehgal, R. Allison, N. Battaglia, J. R. Bond, E. Calabrese, J. Caligiuri, K. Coughlin, D. Crichton, R. Datta *et al.*, "Evidence of lensing of the cosmic microwave background by dark matter halos," *Physical review letters*, vol. 114, no. 15, p. 151302, 2015.

[39] A. Sommerfeld, "Über die beugung und bremsung der elektronen," *Annalen der Physik*, vol. 403, no. 3, pp. 257–330, 1931.

[40] J. F. Navarro, C. S. Frenk, and S. D. White, "A universal density profile from hier-

archical clustering," *The Astrophysical Journal*, vol. 490, no. 2, p. 493, 1997.

[41] K. Freese, P. Gondolo, J. Sellwood, and D. Spolyar, "Dark matter densities during the formation of the first stars and in dark stars," *The Astrophysical Journal*, vol. 693, no. 2, p. 1563, 2009.

[42] M. Ackermann, M. Ajello, W. Atwood, L. Baldini, G. Barbiellini, D. Bastieri, K. Bechtol, R. Bellazzini, R. Blandford, E. Bloom *et al.*, "Constraints on the galactic halo dark matter from fermi-lat diffuse measurements," *The Astrophysical Journal*, vol. 761, no. 2, p. 91, 2012.

[43] B. Anderson, M. Kuhlen, J. Diemand, R. P. Johnson, and P. Madau, "Fermi-lat sensitivity to dark matter annihilation in via lactea ii substructure," *The Astrophysical Journal*, vol. 718, no. 2, p. 899, 2010.

[44] A. Berlin and D. Hooper, "Stringent constraints on the dark matter annihilation cross section from subhalo searches with the fermi gamma-ray space telescope," *Physical Review D*, vol. 89, no. 1, p. 016014, 2014.

[45] J. L. Feng, M. Kaplinghat, and H.-B. Yu, "Sommerfeld enhancements for thermal relic dark matter," *Physical Review D*, vol. 82, no. 8, p. 083525, 2010.

[46] J. Diemand, M. Kuhlen, and P. Madau, "Dark matter substructure and gamma-ray annihilation in the milky way halo," *The Astrophysical Journal*, vol. 657, no. 1, p. 262, 2007.

[47] ——, "Formation and evolution of galaxy dark matter halos and their substructure,"

The Astrophysical Journal, vol. 667, no. 2, p. 859, 2007.

[48] M. Kuhlen, J. Diemand, and P. Madau, "The shapes, orientation, and alignment of galactic dark matter subhalos," *The Astrophysical Journal*, vol. 671, no. 2, p. 1135, 2007.

[49] K. M. Gorski, E. Hivon, A. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann, "Healpix: a framework for high-resolution discretization and fast analysis of data distributed on the sphere," *The Astrophysical Journal*, vol. 622, no. 2, p. 759, 2005.

[50] L. Yang and A. Szalay, "A gpu-based visualization method for computing dark matter annihilation signal," in *Astronomical Data Analysis Software and Systems XXII*, vol. 475, 2013, p. 73.

[51] M. Aguilar, G. Alberti, B. Alpat, A. Alvino, G. Ambrosi, K. Andeen, H. Anderhub, L. Arruda, P. Azzarello, A. Bachlechner *et al.*, "First result from the alpha magnetic spectrometer on the international space station: precision measurement of the positron fraction in primary cosmic rays of 0.5–350 gev," *Physical Review Letters*, vol. 110, no. 14, p. 141102, 2013.

[52] M. Lattanzi and J. Silk, "Can the wimp annihilation boost factor be boosted by the sommerfeld enhancement?" *Physical Review D*, vol. 79, no. 8, p. 083523, 2009.

[53] N. Arkani-Hamed, D. P. Finkbeiner, T. R. Slatyer, and N. Weiner, "A theory of dark matter," *Physical Review D*, vol. 79, no. 1, p. 015014, 2009.

[54] M. Kuhlen, P. Madau, and J. Silk, "Exploring dark matter with milky way substructure," *Science*, vol. 325, no. 5943, pp. 970–973, 2009.

[55] H. Goldberg, "Constraint on the photino mass from cosmology," *Physical Review Letters*, vol. 50, no. 19, p. 1419, 1983.

[56] R. Essig, E. Kuflik, S. D. McDermott, T. Volansky, and K. M. Zurek, "Constraining light dark matter with diffuse x-ray and gamma-ray observations," *arXiv preprint arXiv:1309.4091*, 2013.

[57] O. Y. Gnedin, D. Ceverino, N. Y. Gnedin, A. A. Klypin, A. V. Kravtsov, R. Levine, D. Nagai, and G. Yepes, "Halo contraction effect in hydrodynamic simulations of galaxy formation," *arXiv preprint arXiv:1108.5736*, 2011.

[58] G. R. Blumenthal, S. Faber, R. Flores, and J. R. Primack, "Contraction of dark matter galactic halos due to baryonic infall," *The Astrophysical Journal*, vol. 301, pp. 27–34, 1986.

[59] O. Y. Gnedin, A. V. Kravtsov, A. A. Klypin, and D. Nagai, "Response of dark matter halos to condensation of baryons: cosmological simulations and improved adiabatic contraction model," *The Astrophysical Journal*, vol. 616, no. 1, p. 16, 2004.

[60] G. A. Gómez-Vargas, M. A. Sánchez-Conde, J.-H. Huh, M. Peiró, F. Prada, A. Morselli, A. Klypin, D. G. Cerdeno, Y. Mambrini, and C. Muñoz, "Constraints on wimp annihilation for contracted dark matter in the inner galaxy with the fermi-lat," *Journal of Cosmology and Astroparticle Physics*, vol. 2013, no. 10, p. 029, 2013.

[61] K. N. Abazajian, "The consistency of fermi-lat observations of the galactic center with a millisecond pulsar population in the central stellar cluster," *Journal of Cosmology and Astroparticle Physics*, vol. 2011, no. 03, p. 010, 2011.

[62] D. Hooper, I. Cholis, T. Linden, J. M. Siegal-Gaskins, and T. R. Slatyer, "Millisecond pulsars cannot account for the inner galaxys gev excess," *Physical Review D*, vol. 88, no. 8, p. 083009, 2013.

[63] A. V. Belikov, M. R. Buckley, and D. Hooper, "Searching for dark matter subhalos in the fermi-lat second source catalog," *Physical Review D*, vol. 86, no. 4, p. 043504, 2012.

[64] C. Gordon and O. Macias, "Dark matter and pulsar model constraints from galactic center fermi-lat gamma-ray observations," *Physical Review D*, vol. 88, no. 8, p. 083521, 2013.

[65] D. Hooper and T. R. Slatyer, "Two emission mechanisms in the fermi bubbles: a possible signal of annihilating dark matter," *Physics of the Dark Universe*, vol. 2, no. 3, pp. 118–138, 2013.

[66] L. F. Yang, M. C. Neyrinck, M. A. Aragón-Calvo, B. Falck, and J. Silk, "Warmth elevating the depths: shallower voids with warm dark matter," *Monthly Notices of the Royal Astronomical Society*, vol. 451, no. 4, pp. 3606–3614, 2015.

[67] P. J. E. Peebles, *The large-scale structure of the universe.* Princeton, N.J., Princeton University Press, 1980. 435 p., 1980.

[68] G. Bertone, D. Hooper, and J. Silk, "Particle dark matter: Evidence, candidates and constraints," *Physics Reports*, vol. 405, no. 5, pp. 279–390, 2005.

[69] R. Agnese, Z. Ahmed, A. J. Anderson, S. Arrenberg, D. Balakishiyeva, R. Basu Thakur, D. A. Bauer, J. Billard, A. Borgland, D. Brandt, P. L. Brink, T. Bruch, R. Bunker, B. Cabrera, D. O. Caldwell, D. G. Cerdeno, H. Chagani, J. Cooley, B. Cornell, C. H. Crewdson, P. Cushman, M. Daal, F. Dejongh, E. do Couto e Silva, T. Doughty, L. Esteban, S. Fallows, E. Figueroa-Feliciano, J. Filippini, J. Fox, M. Fritts, G. L. Godfrey, S. R. Golwala, J. Hall, R. H. Harris, S. A. Hertel, T. Hofer, D. Holmgren, L. Hsu, M. E. Huber, A. Jastram, O. Kamaev, B. Kara, M. H. Kelsey, A. Kennedy, P. Kim, M. Kiveni, K. Koch, M. Kos, S. W. Leman, B. Loer, E. Lopez Asamar, R. Mahapatra, V. Mandic, C. Martinez, K. A. McCarthy, N. Mirabolfathi, R. A. Moffatt, D. C. Moore, P. Nadeau, R. H. Nelson, K. Page, R. Partridge, M. Pepin, A. Phipps, K. Prasad, M. Pyle, H. Qiu, W. Rau, P. Redl, A. Reisetter, Y. Ricci, T. Saab, B. Sadoulet, J. Sander, K. Schneck, R. W. Schnee, S. Scorza, B. Serfass, B. Shank, D. Speller, K. M. Sundqvist, A. N. Villano, B. Welliver, D. H. Wright, S. Yellin, J. J. Yen, J. Yoo, B. A. Young, and J. Zhang, "Silicon detector dark matter results from the final exposure of cdms ii," *Phys. Rev. Lett.*, vol. 111, p. 251301, Dec 2013. [Online]. Available: http://link.aps.org/doi/10.1103/PhysRevLett.111.251301

[70] E. Aprile, "The xenon1t dark matter search experiment," in *Sources and Detection of Dark Matter and Dark Energy in the Universe*. Springer, 2013, pp. 93–96.

[71] M. Ackermann, A. Albert, B. Anderson, L. Baldini, J. Ballet, G. Barbiellini, D. Bastieri, K. Bechtol, R. Bellazzini, E. Bissaldi, E. D. Bloom, E. Bonamente, A. Bouvier, T. J. Brandt, J. Bregeon, M. Brigida, P. Bruel, R. Buehler, S. Buson, G. A. Caliandro, R. A. Cameron, M. Caragiulo, P. A. Caraveo, C. Cecchi, E. Charles, A. Chekhtman, J. Chiang, S. Ciprini, R. Claus, J. Cohen-Tanugi, J. Conrad, F. D'Ammando, A. de Angelis, C. D. Dermer, S. W. Digel, E. do Couto e Silva, P. S. Drell, A. Drlica-Wagner, R. Essig, C. Favuzzi, E. C. Ferrara, A. Franckowiak, Y. Fukazawa, S. Funk, P. Fusco, F. Gargano, D. Gasparrini, N. Giglietto, M. Giroletti, G. Godfrey, G. A. Gomez-Vargas, I. A. Grenier, S. Guiriec, M. Gustafsson, M. Hayashida, E. Hays, J. Hewitt, R. E. Hughes, T. Jogler, T. Kamae, J. Knödlseder, D. Kocevski, M. Kuss, S. Larsson, L. Latronico, M. Llena Garde, F. Longo, F. Loparco, M. N. Lovellette, P. Lubrano, G. Martinez, M. Mayer, M. N. Mazziotta, P. F. Michelson, W. Mitthumsiri, T. Mizuno, A. A. Moiseev, M. E. Monzani, A. Morselli, I. V. Moskalenko, S. Murgia, R. Nemmen, E. Nuss, T. Ohsugi, E. Orlando, J. F. Ormes, J. S. Perkins, F. Piron, G. Pivato, T. A. Porter, S. Rainò, R. Rando, M. Razzano, S. Razzaque, A. Reimer, O. Reimer, S. Ritz, M. Sánchez-Conde, N. Sehgal, C. Sgrò, E. J. Siskind, P. Spinelli, L. Strigari, D. J. Suson, H. Tajima, H. Takahashi, J. B. Thayer, L. Tibaldo, M. Tinivella, D. F. Torres, Y. Uchiyama, T. L. Usher, J. Vandenbroucke, G. Vianello, V. Vitale, M. Werner, B. L. Winer, K. S. Wood, M. Wood, G. Zaharijas, S. Zimmer, and Fermi-LAT Collaboration, "Dark matter constraints from observations of 25 MilkyÂ Way satellite galaxies with the Fermi Large Area Telescope," *The Physical Review D*, vol. 89, no. 4,

p. 042001, Feb. 2014.

[72] B. Moore, T. Quinn, F. Governato, J. Stadel, and G. Lake, "Cold collapse and the core catastrophe," *Monthly Notices of the Royal Astronomical Society*, vol. 310, no. 4, pp. 1147–1152, 1999.

[73] M. L. Mateo, "Dwarf Galaxies of the Local Group," *Annual Review of Astronomy and Astrophysics*, vol. 36, pp. 435–506, 1998.

[74] S. Ghigna, B. Moore, F. Governato, G. Lake, T. Quinn, and J. Stadel, "Density profiles and substructure of dark matter halos: converging results at ultra-high numerical resolution," *The Astrophysical Journal*, vol. 544, no. 2, p. 616, 2000.

[75] P. J. E. Peebles, "The Void Phenomenon," *The Astrophysical Journal*, vol. 557, pp. 495–504, Aug. 2001.

[76] M. Hoeft, G. Yepes, S. Gottlöber, and V. Springel, "Dwarf galaxies in voids: suppressing star formation with photoheating," *Monthly Notices of the Royal Astronomical Society*, vol. 371, no. 1, pp. 401–414, 2006.

[77] R. Schaeffer and J. Silk, "Cold, warm, or hot dark matter-biased galaxy formation and pancakes," *The Astrophysical Journal*, vol. 332, pp. 1–16, 1988.

[78] T. Moroi, H. Murayama, and M. Yamaguchi, "Cosmological constraints on the light stable gravitino," *Physics Letters B*, vol. 303, no. 3, pp. 289–294, 1993.

[79] A. Boyarsky, J. Lesgourgues, O. Ruchayskiy, and M. Viel, "Lyman-α constraints

on warm and on warm-plus-cold dark matter models," *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 05, p. 012, 2009.

[80] P. Bode, J. P. Ostriker, and N. Turok, "Halo formation in warm dark matter models," *The Astrophysical Journal*, vol. 556, no. 1, p. 93, 2001.

[81] M. Viel, G. D. Becker, J. S. Bolton, and M. G. Haehnelt, "Warm dark matter as a solution to the small scale crisis: New constraints from high redshift Lyman-α forest data," *The Physical Review D*, vol. 88, no. 4, p. 043502, Aug. 2013.

[82] M. Miranda and A. V. Macciò, "Constraining warm dark matter using qso gravitational lensing," *Monthly Notices of the Royal Astronomical Society*, vol. 382, no. 3, pp. 1225–1232, 2007.

[83] A. Schneider, R. E. Smith, A. V. Macciò, and B. Moore, "Non-linear evolution of cosmological structures in warm dark matter models," *Monthly Notices of the Royal Astronomical Society*, vol. 424, pp. 684–698, Jul. 2012.

[84] D. S. Reed, A. Schneider, R. E. Smith, D. Potter, J. Stadel, and B. Moore, "The same with less: The cosmic web of warm versus cold dark matter dwarf galaxies," *ArXiv e-prints*, Oct. 2014.

[85] L. Gao and T. Theuns, "Lighting the Universe with Filaments," *Science*, vol. 317, pp. 1527–, Sep. 2007.

[86] L. Gao, T. Theuns, and V. Springel, "Star forming filaments in warm dark models," *ArXiv e-prints*, Mar. 2014.

[87] M. C. Neyrinck, "Low primordial information content in the Milky Way with warm dark matter," *ArXiv e-prints*, Aug. 2014.

[88] B. S. Ryden, "Measuring Q 0 from the Distortion of Voids in Redshift Space," *The Astrophysical Journal*, vol. 452, p. 25, Oct. 1995.

[89] G. Lavaux and B. D. Wandelt, "Precision Cosmography with Stacked Voids," *The Astrophysical Journal*, vol. 754, p. 109, Aug. 2012.

[90] J. Clampitt, Y.-C. Cai, and B. Li, "Voids in modified gravity: excursion set predictions," *Monthly Notices of the Royal Astronomical Society*, p. stt219, 2013.

[91] B. Li, G.-B. Zhao, and K. Koyama, "Haloes and voids in f (r) gravity," *Monthly Notices of the Royal Astronomical Society*, vol. 421, no. 4, pp. 3481–3487, 2012.

[92] M. Kamionkowski, L. Verde, and R. Jimenez, "The void abundance with non-gaussian primordial perturbations," *Journal of Cosmology and Astroparticle Physics*, vol. 2009, no. 01, p. 010, 2009.

[93] A. V. Tikhonov, S. Gottlöber, G. Yepes, and Y. Hoffman, "The sizes of minivoids in the local universe: an argument in favour of a warm dark matter model?" *Monthly Notices of the Royal Astronomical Society*, vol. 399, no. 3, pp. 1611–1621, 2009.

[94] J. Clampitt and B. Jain, "Lensing Measurements of the Mass Distribution in SDSS Voids," *ArXiv e-prints*, Apr. 2014.

[95] N. Hamaus, P. M. Sutter, and B. D. Wandelt, "Universal Density Profile for Cosmic Voids," *Physical Review Letters*, vol. 112, no. 25, p. 251302, Jun. 2014.

[96] M. C. Neyrinck, "ZOBOV: a parameter-free void-finding algorithm," *Monthly Notices of the Royal Astronomical Society*, vol. 386, pp. 2101–2109, Jun. 2008.

[97] B. L. Falck, M. C. Neyrinck, and A. S. Szalay, "ORIGAMI: Delineating Halos Using Phase-space Folds," *The Astrophysical Journal*, vol. 754, p. 126, Aug. 2012.

[98] V. Springel, "The cosmological simulation code GADGET-2," *Monthly Notices of the Royal Astronomical Society*, vol. 364, pp. 1105–1134, Dec. 2005.

[99] Y. B. Zel'dovich, "Gravitational instability: An approximate theory for large density perturbations." *Astronomy & Astrophysics* , vol. 5, pp. 84–89, Mar. 1970.

[100] A. Lewis, A. Challinor, and A. Lasenby, "Efficient Computation of Cosmic Microwave Background Anisotropies in Closed Friedmann-Robertson-Walker Models," *The Astrophysical Journal*, vol. 538, pp. 473–476, Aug. 2000, http://www.camb.info/.

[101] F. Bernardeau, "The nonlinear evolution of rare events," *The Astrophysical Journal*, vol. 427, pp. 51–71, May 1994.

[102] Z. A. M. Protogeros and R. J. Scherrer, "Local Lagrangian approximations for the evolution of the density distribution function in large-scale structure," *Monthly Notices of the Royal Astronomical Society*, vol. 284, pp. 425–438, Jan. 1997.

[103] M. C. Neyrinck, "Quantifying distortions of the Lagrangian dark-matter mesh in cosmology," *Monthly Notices of the Royal Astronomical Society*, vol. 428, pp. 141–153, Jan. 2013.

[104] R. E. Angulo, O. Hahn, and T. Abel, "The warm dark matter halo mass function

below the cut-off scale," *Monthly Notices of the Royal Astronomical Society*, vol. 434, no. 4, pp. 3337–3347, 2013.

[105] J. Wang and S. D. M. White, "Discreteness effects in simulations of hot/warm dark matter," *Monthly Notices of the Royal Astronomical Society*, vol. 380, pp. 93–103, Sep. 2007.

[106] V. Icke, "Voids and filaments," *Monthly Notices of the Royal Astronomical Society*, vol. 206, pp. 1P–3P, Jan. 1984.

[107] T. Abel, O. Hahn, and R. Kaehler, "Tracing the dark matter sheet in phase space," *Monthly Notices of the Royal Astronomical Society*, vol. 427, pp. 61–76, Nov. 2012.

[108] W. E. Schaap and R. van de Weygaert, "Continuous fields and discrete samples: reconstruction through Delaunay tessellations," *Astronomy & Astrophysics* , vol. 363, pp. L29–L32, Nov. 2000.

[109] R. van de Weygaert and W. Schaap, "The Cosmic Web: Geometric Analysis," in *Data Analysis in Cosmology*, ser. Lecture Notes in Physics, Berlin Springer Verlag, V. J. Martínez, E. Saar, E. Martínez-González, and M.-J. Pons-Bordería, Eds., vol. 665, 2009, pp. 291–413.

[110] B. Falck and M. C. Neyrinck, "The Persistent Percolation of Single-Stream Voids," *Monthly Notices of the Royal Astronomical Society, in press*, Oct. 2015.

[111] J. Carron, "On the Incompleteness of the Moment and Correlation Function Hierarchy

as Probes of the Lognormal Field," *The Astrophysical Journal*, vol. 738, p. 86, Sep. 2011.

[112] J. Carron and M. C. Neyrinck, "On the Inadequacy of N-point Correlation Functions to Describe Nonlinear Cosmological Fields: Explicit Examples and Connection to Simulations," *The Astrophysical Journal*, vol. 750, p. 28, May 2012.

[113] M. C. Neyrinck, "Transformationally decoupling clustering and tracer bias," *ArXiv e-prints*, Jul. 2014.

[114] J. C. Hill, B. D. Sherwin, K. M. Smith, G. E. Addison, N. Battaglia, E. S. Battistelli, J. R. Bond, E. Calabrese, M. J. Devlin, J. Dunkley, R. Dunner, T. Essinger-Hileman, M. B. Gralla, A. Hajian, M. Hasselfield, A. D. Hincks, R. Hlozek, J. P. Hughes, A. Kosowsky, T. Louis, D. Marsden, K. Moodley, M. D. Niemack, L. A. Page, B. Partridge, B. Schmitt, N. Sehgal, J. L. Sievers, D. N. Spergel, S. T. Staggs, D. S. Swetz, R. Thornton, H. Trac, and E. J. Wollack, "The Atacama Cosmology Telescope: A Measurement of the Thermal Sunyaev-Zel'dovich One-Point PDF," *ArXiv e-prints*, Nov. 2014.

[115] V. Icke and R. van de Weygaert, "The galaxy distribution as a Voronoi foam," *The Quarterly Journal of the Royal Astronomical Society*, vol. 32, pp. 85–112, Jun. 1991.

[116] M. C. Neyrinck, "Cosmological Origami: Properties of Cosmic-Web Components when a Non-Stretchy Dark-Matter Sheet Folds," *submitted for refereeing to the Proceedings of the 6th International Meeting on Origami in Science, Mathematics, and Education*, Aug. 2014.

[117] M. A. Aragon-Calvo and A. S. Szalay, "The hierarchical structure and dynamics of voids," *Monthly Notices of the Royal Astronomical Society*, vol. 428, pp. 3409–3424, Feb. 2013.

[118] M. A. Aragon-Calvo, R. van de Weygaert, P. A. Araya-Melo, E. Platen, and A. S. Szalay, "Unfolding the hierarchy of voids," *Monthly Notices of the Royal Astronomical Society*, vol. 404, pp. L89–L93, May 2010.

[119] A. Pisani, G. Lavaux, P. M. Sutter, and B. D. Wandelt, "Real-space density profile reconstruction of stacked voids," *Monthly Notices of the Royal Astronomical Society*, vol. 443, pp. 3238–3250, Oct. 2014.

[120] S. Nadathur, S. Hotchkiss, J. M. Diego, I. T. Iliev, S. Gottlöber, W. A. Watson, and G. Yepes, "Self-similarity and universality of void density profiles in simulation and SDSS data," *ArXiv e-prints*, Jul. 2014.

[121] E. Ricciardelli, V. Quilis, and J. Varela, "On the universality of void density profiles," *Monthly Notices of the Royal Astronomical Society*, vol. 440, pp. 601–609, May 2014.

[122] E. Jennings, Y. Li, and W. Hu, "The abundance of voids and the excursion set formalism," *Monthly Notices of the Royal Astronomical Society*, vol. 434, pp. 2167–2181, Sep. 2013.

[123] B. R. Granett, M. C. Neyrinck, and I. Szapudi, "An Imprint of Superstructures on the Microwave Background due to the Integrated Sachs-Wolfe Effect," *The Astrophysical Journal Letters*, vol. 683, pp. L99–L102, Aug. 2008.

[124] N. Tejos, S. L. Morris, N. H. M. Crighton, T. Theuns, G. Altay, and C. W. Finn, "Large-scale structure in absorption: gas within and around galaxy voids," *Monthly Notices of the Royal Astronomical Society*, vol. 425, pp. 245–260, Sep. 2012.

[125] M. C. Neyrinck and L. F. Yang, "Ringing the initial universe: the response of over-density and transformed-density power spectra to initial spikes," *Monthly Notices of the Royal Astronomical Society*, vol. 433, no. 2, pp. 1628–1633, 2013.

[126] A. Meiksin and M. White, "The growth of correlations in the matter power spectrum," *Monthly Notices of the Royal Astronomical Society*, vol. 308, no. 4, pp. 1179–1184, 1999.

[127] C. D. Rimes and A. J. Hamilton, "Information content of the non-linear power spectrum: the effect of beat-coupling to large scales," *Monthly Notices of the Royal Astronomical Society*, vol. 371, no. 3, pp. 1205–1215, 2006.

[128] M. C. Neyrinck, I. Szapudi, and C. D. Rimes, "Information content in the halo-model dark-matter power spectrum," *Monthly Notices of the Royal Astronomical Society: Letters*, vol. 370, no. 1, pp. L66–L70, 2006.

[129] R. Takahashi, N. Yoshida, M. Takada, T. Matsubara, N. Sugiyama, I. Kayo, A. J. Nishizawa, T. Nishimichi, S. Saito, and A. Taruya, "Simulations of Baryon Acoustic Oscillations. II. Covariance Matrix of the Matter Power Spectrum," *The Astrophysical Journal*, vol. 700, pp. 479–490, Jul. 2009.

[130] A. Kiessling, A. N. Taylor, and A. F. Heavens, "Simulating the Effect of Non-Linear

Mode-Coupling in Cosmological Parameter Estimation," *Mon. Not. Roy. Astron. Soc.*, vol. 416, p. 1045, 2011.

[131] M. C. Neyrinck, I. Szapudi, and A. S. Szalay, "Rejuvenating the Matter Power Spectrum: Restoring Information with a Logarithmic Density Mapping," *The Astrophysical Journal Letters*, vol. 698, pp. L90–L93, Jun. 2009.

[132] H.-J. Seo, M. Sato, S. Dodelson, B. Jain, and M. Takada, "Re-capturing Cosmic Information," *The Astrophysical Journal Letters*, vol. 729, p. L11, Mar. 2011.

[133] M. C. Neyrinck, "Rejuvenating the Matter Power Spectrum. III. The Cosmology Sensitivity of Gaussianized Power Spectra," *The Astrophysical Journal*, vol. 742, p. 91, Dec. 2011.

[134] Y. Yu, P. Zhang, W. Lin, W. Cui, and J. N. Fry, "Gaussianizing the non-Gaussian lensing convergence field: The performance of the Gaussianization," *The Physical Review D*, vol. 84, no. 2, p. 023523, Jul. 2011.

[135] M. Crocce and R. Scoccimarro, "Memory of initial conditions in gravitational clustering," *The Physical Review D*, vol. 73, no. 6, pp. 063 520–+, Mar. 2006.

[136] N. McCullagh, M. C. Neyrinck, I. Szapudi, and A. S. Szalay, "Removing Baryon-acoustic-oscillation Peak Shifts with Local Density Transforms," *The Astrophysical Journal Letters*, vol. 763, p. L14, Jan. 2013.

[137] P. Coles and B. Jones, "A lognormal model for the cosmological mass distribution," *Monthly Notices of the Royal Astronomical Society*, vol. 248, pp. 1–13, Jan. 1991.

[138] D. H. Weinberg, "Reconstructing primordial density fluctuations. I - Method," *Monthly Notices of the Royal Astronomical Society*, vol. 254, pp. 315–342, Jan. 1992.

[139] R. A. C. Croft, D. H. Weinberg, N. Katz, and L. Hernquist, "Recovery of the Power Spectrum of Mass Fluctuations from Observations of the Ly alpha Forest," *The Astrophysical Journal*, vol. 495, p. 44, Mar. 1998.

[140] R. A. C. Croft, D. H. Weinberg, M. Bolte, S. Burles, L. Hernquist, N. Katz, D. Kirkman, and D. Tytler, "Toward a Precise Measurement of Matter Clustering: Lyα Forest Data at Redshifts 2-4," *The Astrophysical Journal*, vol. 581, pp. 20–52, Dec. 2002.

[141] S. F. Shandarin and A. L. Melott, "Minimal spectrum of long-wave perturbations - Is linear growth correct only in the nonlinear regime?" *The Astrophysical Journal*, vol. 364, pp. 396–399, Dec. 1990.

[142] B. Little, D. H. Weinberg, and C. Park, "Primordial fluctuations and non-linear structure," *Monthly Notices of the Royal Astronomical Society*, vol. 253, pp. 295–306, Nov. 1991.

[143] I. Suhhonenko, J. Einasto, L. J. Liivamägi, E. Saar, M. Einasto, G. Hütsi, V. Müller, A. A. Starobinsky, E. Tago, and E. Tempel, "The cosmic web for density perturbations of various scales," *Astronomy & Astrophysics* , vol. 531, p. A149, Jul. 2011.

[144] M. A. Aragon-Calvo, "The MIP Ensemble Simulation: Local Ensemble Statistics in the Cosmic Web," *Mon. Not. Roy. Astron. Soc.*, vol. 455, no. 1, pp. 438–448, 2016.

[145] S. Shandarin, S. Habib, and K. Heitmann, "Cosmic web, multistream flows, and tessellations," *The Physical Review D*, vol. 85, no. 8, p. 083005, Apr. 2012.

[146] M. C. Neyrinck, "Origami constraints on the initial-conditions arrangement of dark-matter caustics and streams," *Monthly Notices of the Royal Astronomical Society*, vol. 427, pp. 494–501, Nov. 2012.

[147] A. J. S. Hamilton, P. Kumar, E. Lu, and A. Matthews, "Reconstructing the primordial spectrum of fluctuations of the universe from the observed nonlinear clustering of galaxies," *The Astrophysical Journal Letters*, vol. 374, pp. L1–L4, Jun. 1991.

[148] J. A. Peacock and S. J. Dodds, "Non-linear evolution of cosmological power spectra," *Monthly Notices of the Royal Astronomical Society*, vol. 280, pp. L19–L26, Jun. 1996.

[149] X. Wang, M. Neyrinck, I. Szapudi, A. Szalay, X. Chen, J. Lesgourgues, A. Riotto, and M. Sloth, "Perturbation Theory of the Cosmological Log-density Field," *The Astrophysical Journal*, vol. 735, p. 32, Jul. 2011.

[150] R. Mohayaee, H. Mathis, S. Colombi, and J. Silk, "Reconstruction of primordial density fields," *Monthly Notices of the Royal Astronomical Society*, vol. 365, pp. 939–959, Jan. 2006.

[151] E. Lawrence, K. Heitmann, M. White, D. Higdon, C. Wagner, S. Habib, and B. Williams, "The Coyote Universe. III. Simulation Suite and Precision Emulator for the Nonlinear Matter Power Spectrum," *The Astrophysical Journal*, vol. 713, pp. 1322–1331, Apr. 2010.

[152] M. A. Aragon-Calvo and L. F. Yang, "The hierarchical nature of the spin alignment of dark matter haloes in filaments," *Monthly Notices of the Royal Astronomical Society: Letters*, p. slu009, 2014.

[153] F. Hoyle, "Problems of cosmical aerodynamics," *Central Air Documents Office, Dayton*, p. 195, 1949.

[154] S. D. M. White, "Angular momentum growth in protogalaxies," *The Astrophysical Journal*, vol. 286, pp. 38–41, Nov. 1984.

[155] N. Kashikawa and S. Okamura, "Spatial orientation of spin vectors of galaxies in the Local Supercluster," *Publications of the Astronomical Society of Japan*, vol. 44, pp. 493–507, Oct. 1992.

[156] J. F. Navarro, M. G. Abadi, and M. Steinmetz, "Tidal Torques and the Orientation of Nearby Disk Galaxies," *The Astrophysical Journal Letters*, vol. 613, pp. L41–L44, Sep. 2004.

[157] S. G. Patiri, A. J. Cuesta, F. Prada, J. Betancort-Rijo, and A. Klypin, "The Alignment of Dark Matter Halos with the Cosmic Web," *The Astrophysical Journal Letters*, vol. 652, pp. L75–L78, Dec. 2006.

[158] I. Trujillo, C. Carretero, and S. G. Patiri, "Detection of the Effect of Cosmological Large-Scale Structure on the Orientation of Galaxies," *The Astrophysical Journal Letters*, vol. 640, pp. L111–L114, Apr. 2006.

[159] J. Lee and P. Erdogdu, "The Alignments of the Galaxy Spins with the Real-Space

Tidal Field Reconstructed from the 2MASS Redshift Survey," *The Astrophysical Journal*, vol. 671, pp. 1248–1255, Dec. 2007.

[160] R. Brunino, I. Trujillo, F. R. Pearce, and P. A. Thomas, "The orientation of galaxy dark matter haloes around cosmic voids," *Monthly Notices of the Royal Astronomical Society*, vol. 375, pp. 184–190, Feb. 2007.

[161] J. Varela, J. Betancort-Rijo, I. Trujillo, and E. Ricciardelli, "The Orientation of Disk Galaxies around Large Cosmic Voids," *The Astrophysical Journal*, vol. 744, p. 82, Jan. 2012.

[162] E. Tempel, R. S. Stoica, and E. Saar, "Evidence for spin alignment of spiral and elliptical/S0 galaxies in filaments," *Monthly Notices of the Royal Astronomical Society*, vol. 428, pp. 1827–1836, Jan. 2013.

[163] M. A. Aragón-Calvo, B. J. T. Jones, R. van de Weygaert, and J. M. van der Hulst, "The multiscale morphology filter: identifying and extracting spatial patterns in the galaxy distribution," *Astronomy & Astrophysics* , vol. 474, pp. 315–338, Oct. 2007.

[164] O. Hahn, C. M. Carollo, C. Porciani, and A. Dekel, "The evolution of dark matter halo properties in clusters, filaments, sheets and voids," *Monthly Notices of the Royal Astronomical Society*, vol. 381, pp. 41–51, Oct. 2007.

[165] Y. Zhang, X. Yang, A. Faltenbacher, V. Springel, W. Lin, and H. Wang, "The Spin and Orientation of Dark Matter Halos Within Cosmic Filaments," *The Astrophysical Journal*, vol. 706, pp. 747–761, Nov. 2009.

[166] S. Codis, C. Pichon, J. Devriendt, A. Slyz, D. Pogosyan, Y. Dubois, and T. Sousbie, "Connecting the cosmic web to the spin of dark haloes: implications for galaxy formation," *Monthly Notices of the Royal Astronomical Society*, vol. 427, pp. 3320–3336, Dec. 2012.

[167] H. E. Trowland, G. F. Lewis, and J. Bland-Hawthorn, "The Cosmic History of the Spin of Dark Matter Halos within the Large-scale Structure," *The Astrophysical Journal*, vol. 762, p. 72, Jan. 2013.

[168] N. I. Libeskind, Y. Hoffman, J. Forero-Romero, S. Gottlöber, A. Knebe, M. Steinmetz, and A. Klypin, "The velocity shear tensor: tracer of halo alignment," *Monthly Notices of the Royal Astronomical Society*, vol. 428, pp. 2489–2499, Jan. 2013.

[169] M. A. A. Calvo, "Morphology and dynamics of the cosmic web," Ph.D. dissertation, University Library of Groningen][Host], 2008.

[170] R. K. Sheth and R. van de Weygaert, "A hierarchy of voids: much ado about nothing," *Monthly Notices of the Royal Astronomical Society*, vol. 350, pp. 517–538, May 2004.

[171] J. Einasto, G. Hütsi, E. Saar, I. Suhhonenko, L. J. Liivamägi, M. Einasto, V. Müller, A. A. Starobinsky, E. Tago, and E. Tempel, "Wavelet analysis of the cosmic web formation," *Astronomy & Astrophysics* , vol. 531, p. A75, Jul. 2011.

[172] B. J. T. Jones, R. van de Weygaert, and M. A. Aragón-Calvo, "Fossil evidence for spin alignment of Sloan Digital Sky Survey galaxies in filaments," *Monthly Notices of the Royal Astronomical Society*, vol. 408, pp. 897–918, Oct. 2010.

[173] C. Pichon and F. Bernardeau, "Vorticity generation in large-scale structure caustics," *Astronomy & Astrophysics* , vol. 343, pp. 663–681, Mar. 1999.

[174] D. Aubert, C. Pichon, and S. Colombi, "The origin and implications of dark matter anisotropic cosmic infall on ˜L_* haloes," *Monthly Notices of the Royal Astronomical Society*, vol. 352, pp. 376–398, Aug. 2004.

[175] T. Sousbie, C. Pichon, S. Colombi, D. Novikov, and D. Pogosyan, "The 3D skeleton: tracing the filamentary structure of the Universe," *Monthly Notices of the Royal Astronomical Society*, vol. 383, pp. 1655–1670, Feb. 2008.

[176] N. I. Libeskind, Y. Hoffman, A. Knebe, M. Steinmetz, S. Gottlöber, O. Metuki, and G. Yepes, "The cosmic web and the orientation of angular momenta," *Monthly Notices of the Royal Astronomical Society*, vol. 421, pp. L137–L141, Mar. 2012.

[177] M. C. Neyrinck, M. A. Aragón-Calvo, D. Jeong, and X. Wang, "A halo bias function measured deeply into voids without stochasticity," *Monthly Notices of the Royal Astronomical Society*, vol. 441, no. 1, pp. 646–655, 2014.

[178] X. Wang, A. Szalay, M. A. Aragón-Calvo, M. C. Neyrinck, and G. L. Eyink, "Kinematic morphology of large-scale structure: evolution from potential to rotational flow," *The Astrophysical Journal*, vol. 793, no. 1, p. 58, 2014.

[179] M. A. Aragón-Calvo, R. van de Weygaert, B. J. T. Jones, and J. M. van der Hulst, "Spin Alignment of Dark Matter Halos in Filaments and Walls," *The Astrophysical Journal Letters*, vol. 655, pp. L5–L8, Jan. 2007.

[180] M. Cautun, R. van de Weygaert, and B. J. T. Jones, "NEXUS: tracing the cosmic web connection," *Monthly Notices of the Royal Astronomical Society*, vol. 429, pp. 1286–1308, Feb. 2013.

[181] L. F. Yang, "An gpu implementation of ltfe estimation," https://github.com/lyang36/dm-density-estimation-plus, 2014.

[182] Z. Liu, N. Ivkin, L. Yang, M. Neyrinck, G. Lemson, A. Szalay, V. Braverman, T. Budavari, R. Burns, and X. Wang, "Streaming algorithms for halo finders," in *e-Science (e-Science), 2015 IEEE 11th International Conference on.* IEEE, 2015, pp. 342–351.

[183] A. Knebe, F. R. Pearce, H. Lux, Y. Ascasibar, P. Behroozi, J. Casado, C. C. Moran, J. Diemand, and K. Dolag, "Structure finding in cosmological simulations: the state of affairs," *Monthly Notices of the Royal Astronomical Society*, vol. 435, pp. 1618–1658, Oct. 2013.

[184] M. Davis, G. Efstathiou, C. S. Frenk, and S. D. M. White, "The evolution of large-scale structure in a universe dominated by cold dark matter," *The Astrophysical Journal*, vol. 292, pp. 371–394, May 1985.

[185] A. Knebe, S. R. Knollmann, S. I. Muldrew, F. R. Pearce, M. A. Aragon-Calvo, Y. Ascasibar, P. S. Behroozi, D. Ceverino, S. Colombi, J. Diemand, and K. Dolag, "Haloes gone MAD: The Halo-Finder Comparison Project," *Monthly Notices of the Royal Astronomical Society*, vol. 415, pp. 2293–2318, Aug. 2011.

[186] N. Alon, Y. Matias, and M. Szegedy, "The space complexity of approximating the

frequency moments," in *Proceedings of the Twenty-eighth Annual ACM Symposium on Theory of Computing*, ser. STOC '96. New York, NY, USA: ACM, 1996, pp. 20–29. [Online]. Available: http://doi.acm.org/10.1145/237814.237823

[187] Y. Zhang, S. Singh, S. Sen, N. Duffield, and C. Lund, "Online identification of hierarchical heavy hitters: Algorithms, evaluation, and applications," in *Proceedings of the 4th ACM SIGCOMM Conference on Internet Measurement*, ser. IMC '04. New York, NY, USA: ACM, 2004, pp. 101–114. [Online]. Available: http://doi.acm.org/10.1145/1028788.1028802

[188] A. Lall, V. Sekar, M. Ogihara, J. Xu, and H. Zhang, "Data streaming algorithms for estimating entropy of network traffic," in *Proceedings of the Joint International Conference on Measurement and Modeling of Computer Systems*, ser. SIGMETRICS '06/Performance '06. New York, NY, USA: ACM, 2006, pp. 145–156. [Online]. Available: http://doi.acm.org/10.1145/1140277.1140295

[189] H. C. Zhao, A. Lall, M. Ogihara, O. Spatscheck, J. Wang, and J. Xu, "A data streaming algorithm for estimating entropies of od flows," in *Proceedings of the 7th ACM SIGCOMM Conference on Internet Measurement*, ser. IMC '07. New York, NY, USA: ACM, 2007, pp. 279–290. [Online]. Available: http://doi.acm.org/10.1145/1298306.1298345

[190] J. Beringer and E. Hüllermeier, "Efficient instance-based learning on data streams," *Intell. Data Anal.*, vol. 11, no. 6, pp. 627–650, Dec. 2007. [Online]. Available: http://dl.acm.org/citation.cfm?id=1368018.1368022

[191] E. Liberty, "Simple and deterministic matrix sketching," in *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining.* ACM, 2013, pp. 581–588.

[192] M. Kontaki, A. N. Papadopoulos, and Y. Manolopoulos, "Continuous trend-based clustering in data streams," in *Proceedings of the 10th International Conference on Data Warehousing and Knowledge Discovery*, ser. DaWaK '08. Berlin, Heidelberg: Springer-Verlag, 2008, pp. 251–262. [Online]. Available: http://dx.doi.org/10.1007/978-3-540-85836-2_24

[193] L. Serir, E. Ramasso, and N. Zerhouni, "Evidential evolving gustafson-kessel algorithm for online data streams partitioning using belief function theory." *Int. J. Approx. Reasoning*, vol. 53, no. 5, pp. 747–768, 2012. [Online]. Available: http://dblp.uni-trier.de/db/journals/ijar/ijar53.html#SerirRZ12

[194] B. Ball, M. Flood, H. Jagadish, J. Langsam, L. Raschid, and P. Wiriyathammabhum, "A flexible and extensible contract aggregation framework (caf) for financial data stream analytics," in *Proceedings of the International Workshop on Data Science for Macro-Modeling.* ACM, 2014, pp. 1–6.

[195] F. Rusu and A. Dobra, "Statistical analysis of sketch estimators," in *Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data*, ser. SIGMOD '07. New York, NY, USA: ACM, 2007, pp. 187–198. [Online]. Available: http://doi.acm.org/10.1145/1247480.1247503

[196] J. Spiegel and N. Polyzotis, "Graph-based synopses for relational selectivity

estimation," in *Proceedings of the 2006 ACM SIGMOD International Conference on Management of Data*, ser. SIGMOD '06. New York, NY, USA: ACM, 2006, pp. 205–216. [Online]. Available: http://doi.acm.org/10.1145/1142473.1142497

[197] M. Charikar, K. Chen, and M. Farach-Colton, "Finding frequent items in data streams," in *Proceedings of the 29th International Colloquium on Automata, Languages and Programming*, ser. ICALP '02. London, UK, UK: Springer-Verlag, 2002, pp. 693–703. [Online]. Available: http://dl.acm.org/citation.cfm?id=646255.684566

[198] V. Braverman and R. Ostrovsky, "Approximating large frequency moments with pick-and-drop sampling," *CoRR*, vol. abs/1212.0202, 2012. [Online]. Available: http://dblp.uni-trier.de/db/journals/corr/corr1212.html#abs-1212-0202

[199] V. Braverman, J. Katzman, C. Seidell, and G. Vorsanger, "An Optimal Algorithm for Large Frequency Moments Using $O(n^{1-2/k})$ Bits," in *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2014)*, ser. Leibniz International Proceedings in Informatics (LIPIcs), K. Jansen, J. D. P. Rolim, N. R. Devanur, and C. Moore, Eds., vol. 28. Dagstuhl, Germany: Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2014, pp. 531–544. [Online]. Available: http://drops.dagstuhl.de/opus/volltexte/2014/4721

[200] P. Indyk and D. Woodruff, "Optimal approximations of the frequency moments of data streams," in *Proceedings of the Thirty-seventh Annual ACM Symposium on Theory of Computing*, ser. STOC '05. New York, NY, USA: ACM, 2005, pp. 202–208. [Online]. Available: http://doi.acm.org/10.1145/1060590.1060621

[201] I. Kayo, A. Taruya, and Y. Suto, "Probability distribution function of cosmological density fluctuations from a gaussian initial condition: Comparison of one-point and two-point lognormal model predictions with n-body simulations," *The Astrophysical Journal*, vol. 561, no. 1, p. 22, 2001. [Online]. Available: http://stacks.iop.org/0004-637X/561/i=1/a=22

[202] R. E. Smith, J. A. Peacock, A. Jenkins, S. D. M. White, C. S. Frenk, F. R. Pearce, P. A. Thomas, G. Efstathiou, and H. M. P. Couchman, "Stable clustering, the halo model and non-linear cosmological power spectra," *Monthly Notices of the Royal Astronomical Society*, vol. 341, pp. 1311–1332, Jun. 2003.

[203] S. Gottlöber and G. Yepes, "Shape, Spin, and Baryon Fraction of Clusters in the MareNostrum Universe," *The Astrophysical Journal*, vol. 664, pp. 117–122, Jul. 2007.

[204] G. Cormode and M. Hadjieleftheriou, "Finding frequent items in data streams," *Proc. VLDB Endow.*, vol. 1, no. 2, pp. 1530–1541, Aug. 2008. [Online]. Available: http://dx.doi.org/10.14778/1454159.1454225

[205] S. R. Knollmann and A. Knebe, "Ahf: Amiga's halo finder," *The Astrophysical Journal Supplement Series*, vol. 182, no. 2, p. 608, 2009.

[206] S. Planelles and V. Quilis, "Asohf: a new adaptive spherical overdensity halo finder," *Astronomy & Astrophysics*, vol. 519, p. A94, 2010.

[207] A. Klypin and J. Holtzman, "Particle-Mesh code for cosmological simulations," *ArXiv Astrophysics e-prints*, Dec. 1997.

[208] M. C. Neyrinck, N. Y. Gnedin, and A. J. S. Hamilton, "VOBOZ: an almost-parameter-free halo-finding algorithm," *Monthly Notices of the Royal Astronomical Society*, vol. 356, pp. 1222–1232, Feb. 2005.

Vita

Lin Yang was born in China. He obtained bachelor's degree from Tsinghua University, Beijing China (2011).